Tilburg University

Empirical analysis of consumer behavior

Huang, Yufeng

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Huang, Y. (2015). *Empirical analysis of consumer behavior*. CentER, Center for Economic Research.

Empirical Analysis of Consumer Behavior

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof.dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 24 juni 2015 om 16.15 uur door

YUFENG HUANG

geboren op 5 oktober 1987 te Anhui, China.

PROMOTOR: prof.dr.ir. B.J.J.A.M. Bronnenberg
COPROMOTOR: dr. T.J. Klein
OVERIGE COMMISSIELEDEN: prof.dr. J.H. Abbring
prof.dr. M.G. Dekimpe
prof.dr. E. Gijsbrechts
prof.dr. J.P. Dubé
prof.dr. F. Verboven

Acknowledgments

This dissertation contains my work as a PhD student at Department of Marketing in Tilburg University. The development of these essays would not be possible, without the immense help, guidance and support from my mentors and friends. And without my friends, my life in Tilburg would not have been so enjoyable. It is my great pleasure to take this opportunity to express my gratitude.

First, above all, I am deeply indebted to my advisors, Bart Bronnenberg and Tobias Klein. They are not only advisors to my projects, who gave me their opinion, feedback and challenge on my research progress. Also, they are great mentors who taught me ways to think of problems, to write and present, and all important details in life of a researcher. It has been a privilege for me to learn from them in the past years.

When I joined the Marketing department from the economics program, Bart encouraged me to read broadly, both into the "state-of-the-art" marketing literature, and into the classical economic theories and thoughts. The process later inspired the idea behind my job market paper. Bart never pushes his student to meet a deadline; rather, he acts as a role model and exemplifies diligence through his own persistence. Besides research, he concerns the well-being of a student to the tiniest detail. He was as stressed out as I was during the job market, and as happy when I was off the market. The immeasurable support from him would no doubt be a motivation to keep me forward.

Tobias encouraged me to initiate my exploration around the area of structural empirical work, during my research master studies. He gave me much freedom to wander around, but was always there whenever I need guidance. With his encouragement and support, I wrote my first simulation code, laid out my first proof of identification, and talked about my research idea on the spot the first time. From the research master to PhD phase, he always patiently gave me detailed and constructive comments, which shaped my vague and dis-organized research ideas into much more organized ones. His way to thinking about applied economic problems greatly inspired my thoughts.

Besides my two advisors, I am very grateful to Jaap Abbring for his guidance and enlightenment. Although he is not “officially” my advisor, I received much of my methodological training from him. I am also thankful for his suggestions to my dissertation.

For my dissertation, I am also indebted to to my committee for their insightful comments, in particular from Marnik Dekimpe and Els Gijsbrechts. This dissertation also benefited from comments and suggestions from many colleagues, friends and seminar/conference participants; and I am thankful for the feedback from J.P. Dubé, Gautam Gowrisankaran, Günter Hitsch, Shridhar Narayanan, Chris Nosko, Carlos Santos, Peter Rossi, Yifan Yu and Yan Xu, and many others.

Going through the job market was a very stressful experience, but I was blessed to be backed up by the support from many people. Besides my advisors and dissertation committee, I benefited from the helpful suggestions from Otilia Boldea, Hannes Datta, Bettina Drepper, George Knox, Rik Pieters and Joachim Vosgerau. I am also very grateful to my colleagues in the Marketing Department for their support.

My 6 years in Tilburg was made colorful by many of my friends, who become an important part of my life: Zhuojiong Gan, Jonne Guyt, Esther Jaspters, Xue Jia, Kristopher Keller, Shuai Kou, Zhengyu Li, Arjen van Lin, Jinghua Lei, Ana Martinovici, Max Nohe, Constant Pieters, Johanna Slot, Benjamin Tanz, Keyan Wang, Yun Wang, Ruixin Wang, Yan Xu, Soulimane Yajjou, and Yifan Yu (obviously this cannot be an exhaustive list).

Finally, I want to express my gratitude to my parents and my girlfriend. Mom and Dad, as I am your only child, you never wanted me to stay half way across the globe, yet you full-heartedly support my decision to pursue my academic career. I am proud to be your son. And to my girlfriend Bowen, thank you for always being interested, supportive and critical to my ideas, and also for sharing the joy and stress in life with me.

Yufeng Huang

May 13, 2015

Contents

Chapter 1

Introduction

This dissertation contains 3 essays in quantitative marketing, devoted to empirically characterizing consumer decision making. The 3 essays address topics in, respectively, 1) consumer skill evolution and their product usage, 2) decision costs in shopping, and 3) a consumer's choice of a distribution channel.

In the first essay, I empirically characterize the way that a consumer's skill of using a product (her product-specific human capital) evolves as she gains experience from using the product, and the role of her human capital on her choice of using the product, as well as switching to a different product. I chose the digital camera industry as empirical context. On a popular photo-sharing website, I find that viewers tend to take closer looks at pictures taken by individuals with more general experience in photography, without knowing that in advance. On the other hand, they tend to avoid clicking on pictures captured by a camera that the individual recently adopted. This suggests that clicks by the viewers are indicative of how good the pictures look, which is in turn, outcome of the photographer's general, and camera-specific skills. I then estimate a dynamic demand model with learning by doing, i.e., taking into account endogenous evolution of the individual's human capital. I find large returns in general human capital in a consumer's tendency to upgrade to a high-end product, as well as very important role in a user's product-specific (importantly, brand-specific) human capital, of which the evolution shapes an individual's increasing loyalty to products or brands.

In Essay 2, we quantify a consumer's cost of consideration – i.e., the cost of processing information on the less-salient product characteristics. Because the consumer has to incur a cost for each product she considers, she has an incentive to avoid considering too many products. Hence, she will pick subsets of products (consideration sets) that prioritize, for example, products that are sold at a discounted price. Therefore, for a consumer who frequently purchases multiple units, a marginal change in price causes a discontinuous jump in her purchase quantity, due to changes in consideration set membership. We then construct and estimate a model of multiple product-quantity choice, with endogenous consideration set formation. We find that consideration is costly: in a given trip, it costs a consumer about the magnitude of her expenditure in each product, to consider it in the first place. Further, we separate the impact of a price change on purchase quantity given consideration, from its impact on shaping the consideration set itself, and find that the majority of demand's response to price comes from changes in the consumer's decision to consider it or not. Finally, we find that feature advertising reduces the consideration cost, and selects consumers with higher price elasticities. Hence, feature advertising is better coupled with a price discount.

The final essay studies how a consumer substitutes between the two retail channels – in-store shopping or the online counterpart – within the same chain store. We exploit exogenous variations in potential travel costs, due to store outlet exit or entry, or consumer house-moving, and find that consumers are elastic to changes in potential distance; and in addition, their substitutability increases over time. This can be explain by that a consumer's perceived probability of receiving an inferior product (hence the term "perceived risk") decreases over time, and the decrease in risk reduces a consumer's need to shop off-line – through which she can verify the product quality before purchase. We exploit exogenous variations in consumer expenditure, from shifters of income, price and purchases of season-specific products, and find strong evidence in support of the theory. We then quantify a structural model of channel and expenditure choice, and find that consumer's perceived risk is reduced by 3/4, when a consumer becomes very experienced. This change corresponds to 3 times as high the online sales. As managerial implications, we emphasize both potential policy instruments that can facilitate channel migration (through experience), as well as the danger of ignoring online and off-line cannibalization effects.

Chapter 2

Learning by Doing and the Demand for Advanced Products

2.1 Introduction

> "I'd get a DSLR based upon my experience level. [...] If your situation is different to mine however. [...] you'll probably be quite happy with a cheaper point and shoot."
>
> – Darren Rowse,[1] *Should you buy a DSLR or Point and Shoot Digital Camera?*

> "Nikon and Canon are as good as each other overall. [...] The differences lie in ergonomics and how well each camera handles, which is what allows you to get your photo – or miss it forever. [...] and I can't for the life of me figure out the menus of the Nikon Coolpix cameras."
>
> – Ken Rockwell,[2] *Nikon vs. Canon*

[1]Extracted from the following URL by March 2014. http://digital-photography-school.com/should-you-buy-a-dslr-or-point-and-shoot-digital-camera

[2]Extracted from the following URL by March 2014. http://www.kenrockwell.com/tech/nikon-vs-canon.htm

The two quotes above demonstrate a widely-held belief among practitioners – that novices and expert consumers demand products of different quality; and among the experts, their demand is specialized, and thus dependent on their previous experience. The role of consumer product experience is not unique to the digital camera industry. For products such as home electronics, sports equipment, entertainment, food and beverages,[3] marketing practitioners have long realized the wide differences in the demand between novices and experts, and have targeted their different needs by developing portfolios of differentiated products.

Despite the importance of experience accumulation in consumer demand, the *quantitative* understanding of this is limited. This is because standard choice data alone confound the returns to experience with alternative explanations, such as changes in tastes or increases in awareness. In the context of choices of digital cameras, this paper utilizes a unique data-set that provides a measure of the returns to consumer experience, and quantifies its role in the demand for entry-level and advanced digital cameras. This allows for a better understanding of the long-run evolution of demand from entry-level to advanced products, and potentially, better quantitative marketing decisions.

In this paper, a consumer of digital cameras cares about her picture quality, which she *learns to produce* through the accumulation of experience. Hence, being able to measure the effect of experience on picture quality, and the effect of increasing picture quality on her demand for advanced products, is key to understanding the effect of experience on demand in this case. For this purpose, I collect individual panel data from pictures displayed on a photo-sharing website, Flickr.com. And I exploit the fact that pictures are sorted by the date of upload, and I compare the number of views among pictures that a consumer uploads at the same time. Since these pictures are displayed together and are likely to be viewed together, the differences in views are more likely to reflect picture quality differences. In addition, I observe variations in *when* and *by which camera* each picture was *taken*, within the same batch of upload, and hence can infer the causal effect of experience and equipment

[3] Alba and Hutchinson (1987) are among the first to conceptualize the role of consumer product experience – "expertise". In two experimental studies, Nam et al. (2012) and Clarkson et al. (2013) document the differences between expert and novice consumers, in their choices of, respectively, digital cameras and food/beverages. Albuquerque and Nevskaya (2012) model a consumer's progressively higher tendency to play video games. Youn et al. (2008) document that beginner climbers tend to choose entry level climbing gears, and will later progress into advanced but specialized products.

on the picture quality.

With up to 10 years of measurement of picture quality per individual, jointly with observations of camera usage and switching,[4] the role of experience accumulation is evident even without a (structural) model. On the one hand, with their experience accumulating, consumers are capable of producing higher quality pictures. On the other hand, after a consumer switches cameras, she cannot *immediately* produce pictures of as high quality as she did with the previous one; and this gap is larger for consumers with more experience. This indicates that not only is the consumer obtaining general experience in photography, she is also accumulating specific knowledge about using the given product.

To quantify the role of experience on demand in this context, I then construct a structural model of a consumer's demand for cameras and choices of product usage. In the model, the quality of the camera that the consumer owns is *complemented* by her ability to use it – her "human capital", which improves with previous experience through learning by doing. Accumulation of experience thus changes the consumer's relative importance of product quality and price, and spurs the demand for advanced products. However, *part of* the consumer's experience is knowledge on operating a specific camera, and cannot be utilized after she switches to another one. The consumer thus faces a key trade-off. On the one hand, learning by doing encourages her to *delay* switching to advanced products, since higher human capital brings higher immediate benefit for using the product. On the other hand, the longer she waits, the more effort she spends on learning non-transferable, camera-specific features; and because of this, the consumer would rather switch to advanced products *early*.

To ensure that alternative explanations are controlled for, I allow for differences across consumers, in their (time-invariant) preferences as well as the way that past history affects their current choices, which captures across-consumer differences in demand and demand evolution. In addition, the initial period differences, both in prior experience and in the choice of the first camera, are also captured by the model. Finally, I also model technology evolution, and the individual's rational expectation on it.

Both the data and the structural estimates find substantial returns to experience in pho-

[4]The latter is inferred from changes in the identity of the cameras.

tography: on average, a consumer with 5 years of experience produce pictures with higher quality, that attracts more than twice as much attention compared to her earlier pictures. In addition, the knowledge she learned (her human capital) in the 5 years contributes to half of the quality increase. In addition, not all her human capital is transferable, if she switches to other cameras: for example, for an average Canon compact camera user with 2 years of experience, only 1 year of her experience is applicable to a Nikon DSLR camera; for one with 5 years of experience, however, only 2 year-equivalent of her experience can be transferred to the new Nikon advanced camera. This means that the attrition in human capital increasingly discourage a consumer from using (and learning) other cameras, especially those from other brands.

To my knowledge, this is the first paper to structurally quantify the role of accumulation in product usage experience on consumer demand. Although the empirical exercise focuses on choices between entry-level and advanced digital cameras, the insight from this paper can be applied to a broad range of industries, such as home electronics, sports equipment, entertainment, and other categories where usage of products requires consumer human capital. In studying consumer human capital evolution, this paper contributes to practical understanding of the evolution of consumer demand through product usage, as well as consumers' gradual tendency to be locked in to products with similar characteristics – such as brands.

As the first contribution, this paper quantifies the returns to experience on consumers' demand for advanced products. Experience accumulation increases a consumer's payoff from product usage, which in turn increases her demand for advanced products. In the data, consumers gradually switch from compact cameras to DSLR cameras, and I find that learning by doing explains 1/3 of this increase in DSLR market share. This implies that the overall stock of, and the growth in, consumer experience have substantial influence on the market demand for advanced products, and thus offers an explanation of the demand-driven innovation hypothesis (Adner and Levinthal, 2001). Supply-side provision of consumer knowledge – such as free product training, stimulating consumer content creation, or designing products that are easy to use – can facilitate the evolution of demand via the increase in consumer human capital.

Part of the consumer experience is product-specific. As the second contribution, this paper finds that an important barrier to knowledge transfer is the differences in product designs across brands, which creates significant brand loyalty that accumulates through experience. As a consumer's product experience accumulates, she becomes less willing to switch to other brands. I find that the interaction between learning by doing and switching cost – that is, the evolution of product specific human capital – largely explains the (lack of) demand for compact cameras. Because a consumer might end up with an advanced camera in the *long run*, product-specific learning eliminates 15-30% *short-run* demand for compact cameras, despite that they should have been the current optimal choice for many consumers. This mechanism of endogenous switching cost is related to the literature of brand loyalty (Dubé, Hitsch and Rossi, 2009; Dubé et al., 2010) and evolution of consumer brand preferences through experience accumulation (Erdem and Keane, 1996; Bronnenberg et al., 2012).

The remainder of the paper is structured as follows. Section 2.2 gives a brief review to the literature related to this study. Section 2.3 describes the data collection process and how I define the key variables – in particular, the identification strategy that allows us to measure picture quality. Section 2.4 then presents model-free evidence that shows the importance of consumer learning by doing, and the role of switching cost. Given the evidence, Section 2.5 outlines an empirical model of experience evolution and consumer choices on purchasing and using cameras. Next, Section 2.6 presents and discusses parameter estimates, implied state evolution, price elasticities and model fit. Section 2.7 then discusses the managerial implications, and Section 2.8 concludes.

2.2 Related literature

This paper can be positioned in the intersection of two literatures. On the one hand, my discussion of learning by doing draws from previous theoretical work on consumer human capital (Becker, 1965; Michael, 1973; Alba and Hutchinson, 1987; Jovanovic and Nyarko, 1996; Ratchford, 2001). Built from the framework in Becker (1965), Michael (1973) and Ratchford (2001) point out that consumer human capital determines their utility from product consumption. With different methodology, Alba and Hutchinson (1987) categorize the dimensions of consumer "expertise", and point out its difference from a consumer's information set. Jovanovic and Nyarko (1996) build a framework where non-forward-looking, Bayesian individuals update their knowledge on product usage from previous usage experience, and this increases their incentives to ascend to higher-quality products. Their framework is applied in Foster and Rosenzweig (1995) in their empirical study of increasing rural labor productivity and the choice of applying a new agricultural technology.[5] Ratchford (2001) constructs a framework for consumer human capital, and points out its implication for life-cycle consumption, brand loyalty (in particular, related to its non-transferability) and the decisions to search. Built on this literature, this paper is the first empirical study using field data to study the effect of consumers' human capital on their product replacement/upgrade decisions.

On the other hand, the consumer demand framework of this paper is derived from the literature on dynamic discrete choice of differentiated products, for example, Melnikov (2000), Song and Chintagunta (2003) and Gowrisankaran and Rysman (2012). In Melnikov (2000) and Song and Chintagunta (2003), since their interest focuses on first-time adoption decisions, they assume away repeated purchases, and hence greatly simplify computation. Gowrisankaran and Rysman (2012) allow for repeated purchases, but impose a dimensionality-reduction assumption on the state space, to ease the computational burden. In my paper, the focus is on re-purchase rather than first-time purchase decisions, and I need to consider endogenous product usage decisions and the corresponding outcome – in this case the picture quality. I also take into account dynamic optimization under differentiated product characteristics. Specifically, to maintain the key feature of evolving consumer human capital as well as accounting for other (high-dimensional) state variables, I impose a dimensionality-reduction assumption that is in spirit of Gowrisankaran and Rysman (2012), but does not require the extra layer in the fixed point algorithm.

The fact that consumer human capital is not perfectly transferable creates a switching cost. The general topic of switching cost relates to the empirical literature on the effect of switching costs on consumer decisions, in grocery shopping (Dubé et al., 2010), pharma-

[5] A previous version of this paper also uses the Jovanovic and Nyarko framework, i.e. to specify a Bayesian updating process for human capital accumulation – human capital as one minus the posterior variance. Applying to this context, their framework produces similar quantitative insights and a good model fit.

ceutical products (Crawford and Shum, 2005), health care (Nosal, 2012), and many other categories. In this literature, there are various explanations to a consumer's lack of willingness to transition across brands – hence "brand loyalty". In this paper, I propose an alternative mechanism: that consumers are brand loyal because it is difficult for their experience to transfer across brands – possibly due to differences in designs. A similar explanation, "skill-based habits", is proposed by Murray and Häubl (2005) in their experimental studies. In addition, I also demonstrate that this has dynamic implications especially for forward-looking consumers.

This paper is also related to the empirical literature on the effect of consumer learning. This literature (Erdem and Keane 1996; Erdem et al. 2005, among others) characterizes the effect of information of product attributes on consumer demand. In this framework, knowledge also endogenously evolves through past purchase experience, but the main effect of such knowledge is on consumers' belief (i.e. their information sets), while in my model, experience is effective on consumers' *ex post* utility from product usage.[6] Empirically speaking, learning on product attributes tends to stop rather quickly,[7] while in the case of learning by doing, I examine changes in consumer choice patterns over the course of up to 10 years.

2.3 Data

2.3.1 Collection

I extract picture level data from Flickr.com – a popular photo sharing website. Flickr started its business in 2000 by Ludicorp, and was acquired by Yahoo! in 2005. The data extraction was implemented between March 2012 and April 2013, until a major change in user-interface took place on Flickr. During the data collection period, pictures (including

[6]The difference also corresponds to the difference in "familiarity" and "expertise" in Alba and Hutchinson (1987). In Nelson (1970), the different explanations are two aspects of his categorization of experience goods: "*After using ..., its price and quality can be combined to give us posterior estimates of the utility of its purchase.*" [Nelson (1970), "Information and Consumer Behavior", p.313].

[7]For example, Dubé et al. (2010) do not find non-stationarity in the choice pattern for products that are not new to the market.

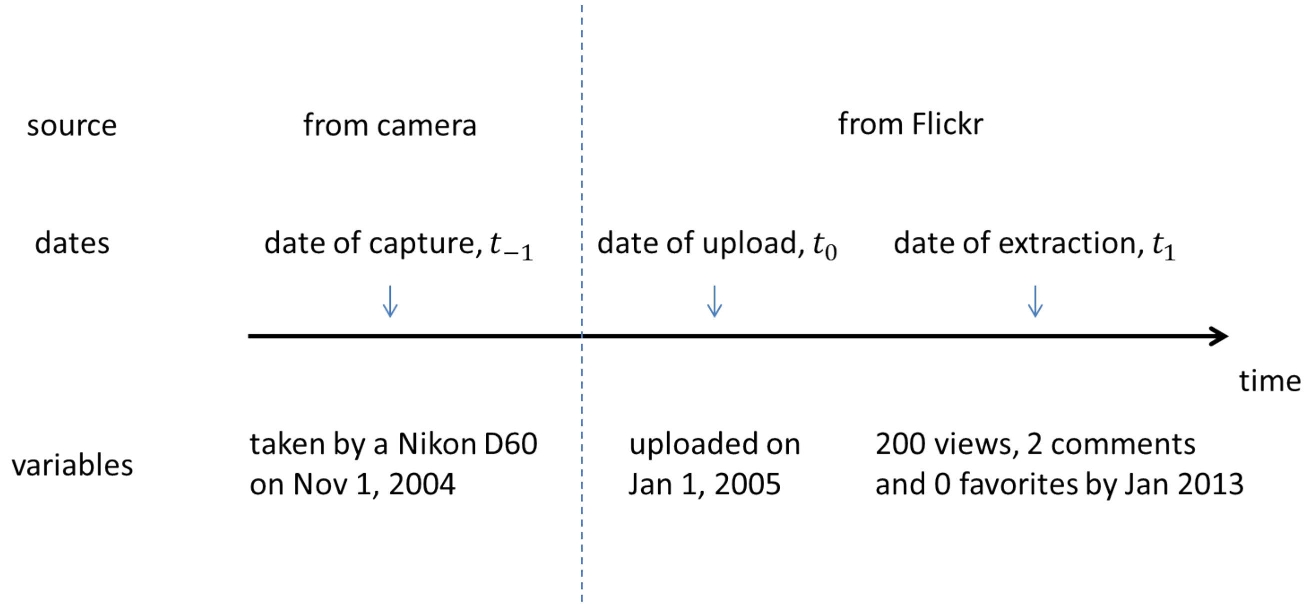


Figure 2.1: Structure of picture-level data from Flickr

Note: This figure depicts the structure of my picture level data, extracted from Flickr.com. The vertical dashed line divides data that are originally recorded by the cameras (Exif data, embedded in each picture), and data that are originally recorded by Flickr. From the camera-recorded data, I collect the camera identity (in this example, Nikon D60) and capture date. From Flickr-recorded data, I collect the upload date, as well as the cumulative views and "favorite" votes from upload to data-extraction. This is done *once* per picture.

their detailed information) were publicly viewable, even without a user account.

Camera-recorded information is embedded in each picture, as *Exif* (exchangeable image file format) data. For the purpose of this paper, those data contain valuable information for camera identity, as well as the date of capture. To complement the Exif data, I also collect information on the date of upload, and the cumulative views and "favorite" votes from the upload till the data-extraction time. Figure 2.1 summarizes the information I get from each picture.

I collect data at two levels. At the picture level, I sort an individual's pictures in order of upload dates, and collect the data *once*, from one in every five pictures. This gives me *cross-sectional* data on picture level information.[8] At the individual level, I collect data on Flickr-summarized monthly picture-taking and uploading records, for each individual.

I also gather a cross-sectional data-set for camera characteristics, and a longitudinal auction price data-set. The camera characteristics data-set is compiled from the Flickr camera

[8]I did not re-visit a picture multiple times, because the time spent on collecting data from each picture is large.

Table 2.1: Sample Selection Criteria

	percentage
Taken by compact camera or DSLR	96.4
Exif data complete	89.6
Taken after year 2000	89.0
All above criteria	75.6
obs.	2777753

Notes: This table reports the sample selection criteria. On the picture level data, I drop the observations on pictures taken by camera formats other than compact camera or DSLR; or with Exif data that lack an indicator of camera model or date of capture; or taken *no later* than January 1, 2000. Altogether, this excludes 24.4% of the sample.

database, DPreview.com, and Cnet.com. In addition, Pixel-peeper.com summarized a long monthly price history for average Ebay auction prices per camera, from 2006 to 2013.

2.3.2 Sample selection and summary statistics

I focus on the users with "pro accounts" (paid accounts), at the time of sampling. These accounts allow users to upload many more pictures than the free accounts. The free account users tend to upload few pictures even over many years. Hence, one might run into the risk of omitting certain camera-switching decisions.[9] On the other hand, however, pro accounts are less representative for the entire market. This paper focuses on within-individual changes, and hence does not require data to be representative over demographics. However, I do not observe account switching histories, and therefore need to assume that account switching is not dependent on learning by doing and camera switching outcomes.

I collect data from all paid account users with a user-name no long than 5 letters/digits. Focusing on shorter user-names gives me users with long histories on Flickr, and usually enables us to observe camera usage in long time spans.[10] On the other hand, it is reasonable to assume that user-names are exogenous to the variables of interest. Sampling one in every 5 pictures gives me close to 2.8 million observations on the picture level. Among these data, I disregard the pictures taken by cell phones, film cameras, camcorders or digital

[9] Flickr offers either a free account – which is imposed a monthly upload capacity as well as a maximum-viewable-pictures restriction (200 pictures in total), or a "pro account" that costs $24.95 (as in 2012) annually. I need to focus on the pro-account users, because limitation to upload capacity will complicate the problem.

[10] The underlying assumption is that the naming strategy is orthogonal to preference and experience. On the other hand, the in-sample duration is not orthogonal to preferences – and hence I do not select on it.

Table 2.2: User Level Data Summary

	Mean	Median	Stdev
months since registered in Flickr	69	74	24
number of contacts at data extraction	94	20	292
total number of pictures	1691	981	1897
number of in-sample pictures	359	203	410
number of cameras ever used in-sample	4	3	4
max views per month, first 10 pic	7	1	57
max views per month, last 10 pic	20	4	193
price of the least expensive camera used	216	157	191
price of the most expensive camera used	1040	762	655
obs.	5499	5499	5499

Notes: The table reports summary statistics from the data. Mean, Median and StDev are the mean, median and standard deviation of the data, respectively. The number of contact is the number of other accounts, who are followed (subscribed) by the given user at the time of data extraction. The number of cameras ever used in-sample is the number of unique camera identities one observes from the user's Exif data. Prices of the least and most expensive cameras are in 2005 US dollars.

media players, or those claimed to be taken prior to year 2000 (which is more likely to be a mistake in the camera date settings), or have incomplete Exif data (in particular when identities of the cameras or the picture taking time are missing). This excludes 24.4% of the picture data – as shown in Table 2.1.

Table 2.2 provides summary statistics for the user level data after sample selection. There are three interesting points to note. First, the median duration of observation for a user is beyond 6 years. This is a long enough period to observe the slow evolution of an individual's photography knowledge. Second, among the 6 years in Flickr, the median user only subscribed to 20 other users. Compared to Facebook users, this shows that (this sample of) Flickr users are not social-network driven.[11] Second, there is a considerable increase in the maximum views per unit time among pictures taken at the beginning of the sample, compared to those taken at the end of the sample; while the views have a larger spread towards the end of the sample. This suggests both an *increase* and a *divergence* in the number of views one's pictures can attract.[12] Third, the median individual has had 3 cameras throughout the 6 years' in-sample period, while there is considerable dispersion in the prices of her camera: the real (Ebay auction) price of her most expensive camera is more than twice of

[11]As a comparison, the median Facebook user has 200 friends, by account of Aaron Smith (extracted in June, 2014, from http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/).

[12]Which might be due to changes in picture quality, or changes in the size of user base of Flickr.com.

the price of her least expensive camera.

2.3.3 (Implied) picture quality

2.3.3.1 Intuition

In this section, I construct a measure for how each picture is received by the (anonymous) viewers on Flickr.com. I call this measure “picture quality”. I then take maximum of this measure for each individual in each period, as a measure of the quality of pictures an individual can produce, using her equipment and her abilities. This measure will be later treated as data in the reduced form analysis and structural estimation.

The basic idea behind the picture quality measure, is that high picture quality is one of the drivers that explains why a picture receives many more views than the others. Especially, among pictures displayed in the same time window, it is likely that they are exposed in front of the same cohort of viewers, and thus their differences in views might better reflect innate quality differences. With this idea, I exploit the variation between the date of capture of a picture, and the date when it was uploaded. Holding the date of upload fixed, differences in views among the pictures should solely reflect differences in their quality – as we are effectively holding the flow of viewers to be the same. My sample consists of more than 158,000 user-months of upload combinations. Among those uploaded in the same month, the first picture was captured 4 months earlier, on average, than the last picture. This gives me ample variation in the capture dates to measure picture quality.

2.3.3.2 Specification and assumptions

Formally, I model the cumulative number of views of picture p captured by individual i, as the accumulation of an underlying viewer-flow process to the photographer i, $flow_{ipt}$, which is by itself multiplicative in the quality of the picture q_{ip}, the overall flow of viewers into Flickr.com ϕ_t, and other observed characteristics of the picture that are not related to quality, z_{ip} (e.g. the topic of the picture, or the order that pictures are displayed might affect their views):

$$views_{ip} = \sum_{t_0 \le t \le t_1} flow_{ipt} \tag{2.1}$$

where

$$flow_{ipt} = \phi_t \exp\left(q_{ip} + z_{ip}\psi\right).$$

Omitting i and p subscripts, I denote t_0 and t_1 to be the calendar dates of upload and data extraction, respectively. Note that t_0 and t_1 are picture specific. The cumulative number of views is the summation of the viewer flow between these two dates. In the viewer flow specification, q_{ip} is the (unobserved) quality of the picture, which is implicitly a function of user experience, camera, and an econometric error.[13]

Take the log of Equation (2.1), we have

$$\log\left(views_{ip}\right) = \Phi_{t_0 t_1} + z_{ip}\psi + q_{ip}, \tag{2.2}$$

noting that $\Phi_{t_0 t_1} = \log\left(\sum_{t_0 \le t \le t_1} \phi_t\right)$ is a time-window-specific fixed effect, that captures the overall cumulative viewer arrival in the time window $[t_0, t_1]$, when the picture was on display.

The specification (2.2) makes two assumptions. First, the upload *timing* decision of picture p is orthogonal to the unobserved quality q_{ip}, up to an individual fixed effect. That is, given an individual's fixed characteristics, she does not time the upload in the order of their quality. For example, this assumption will be violated if an individual decides to upload good pictures first, and the not-so-good (but still good enough that she would upload) pictures in later batches. If this is the case, the upload time t_0, and hence $\Phi_{t_0 t_1}$, will be correlated with q_{ip}. To address this concern, we verify that more than 3/4 of all pictures are uploaded in the immediate next batch, which implies that the (infrequent) upload decision might be driven by other time costs. In addition, as shown in Figure 2.14 in the Supporting Material, the views of the delayed-uploaded pictures are not systematically different from others in its own batch, which suggests that their quality is not selectively different.

As the second assumption, I impose that the cumulative viewer base, $\Phi_{t_0 t_1} = \log\left(\sum_{t_0 \le t \le t_1} \phi_t\right)$, is the same for all individuals who have pictures displayed in the given interval (up to in-

[13] One might alternatively interpret this as a noisy measure of picture quality.

dividual fixed effects). This assumption would be violated if for some individuals, their viewer base *increase* faster, and therefore they have higher $\Phi_{t_0 t_1}$ only in later time intervals (i.e. when t_0 is larger). If this is the case, we will over-state the trend in q_{ip} (which we attribute to learning by doing). To address this concern, in Supporting Material 2.9.1, I document that a reduced-form estimate of the trend in q_{ip} is robust to this assumption.

2.3.3.3 Implementation

With these two assumptions, I estimate Equation (2.2) by ordinary least squares, controlling for combinations of picture upload month and data extraction month ($\Phi_{t_0 t_1}$), as well as individual fixed effects (contained in q_{ip}). I also include the following control variables in z_{ip}: 1) the topic of the pictures, as captured by *tag* fixed effects, 2) the number of pictures uploaded in the same batch, 3) the order of the focal picture in the upload batch, and 4) months since a user was registered on Flickr (as a proxy of the accumulation of friends networks). All control variables are coded as indicator variables, allowing the specification to be as flexible as possible. I take the projected individual fixed effect plus the residual term, as a proxy of picture quality.[14] This gives a measure of quality for each individual picture p, captured by i.

Next, we need to obtain a systematic measure of what an individual can produce (using her camera and human capital) at a given point in time. To this end, because uploading each picture can be costly both in time and in storage space, the observe quality distribution (distribution of the implied individual picture quality, $\hat{q}_{ip}$) might be heavily selected. To address this concern, I assume that there exists an upper bound of picture quality, which represents what an individual can produce using camera j at time t. In addition, she takes many pictures and uploads the best few, so that the best among the uploaded pictures represents this (up to a measurement error). With this idea in mind, I take maximum over all

[14]The reason I consider individual fixed effects as systematic across-individual difference in quality rather than other factors such as being popular on Flickr, is because we can trace every user to her starting point in Flickr, but not to her initial experience in photography. Therefore, it is much more plausible to think of heterogeneity in initial conditions as heterogeneity in skills. As a robustness check, leaving out the individual fixed effect does not qualitatively change the (reduced form and structural) estimates.

implied individual picture quality, $\hat{q}_{ip}$, produced in a given month t:

$$Q_{ijt} = \max_{p \in t} \hat{q}_{ip}$$

where we abuse the notation $p \in t$ to denote pictures produced in time t (by i with camera j). Alternatively, we can also assume away potential selection, and use the average (or other moments of) picture quality to measure human capital. In Section 2.4.4, I present reduced-form evidence for the presence of switching cost, using the alternative quality measures.[15]

2.3.3.4 Summary statistics for picture quality

Table 2.3 summarizes the maximum picture quality in each picture-taking month, which characterizes the quality of pictures that an individual *can* produce. One can immediately spot the following patterns.

First, with accumulating years of experience, the individual *can* produce increasingly higher picture quality, up to a point where knowledge has been saturated, and the change in picture quality is statistically negligible. In other words, there is a clear pattern of learning with decreasing speed.

Second, using a small sub-sample with non-zero favorite-votes data,[16] one can cross-check whether the developed measure of picture quality is reasonable. I find that the correlation between maximum picture quality and maximum rating (if nonzero) is around 60%, which justifies that the maximum quality is a reasonable measure of the outcome of picture taking.[17]

---

[15] In the previous version, I also structurally estimate the model using mean quality, and obtain similar qualitative results.

[16] The share of individual-monthly observations where *at least one* picture has received *at least one* favorite vote is 15%.

[17] A potential concern is that taking more pictures in a given month will drive up the maximum quality and rating. We examine whether the correlation is driven by the number of pictures, by estimating the maximum quality and rating on a flexible function of the number of pictures, and the number of pictures with positive ratings, respectively. One finds that the correlation is high even adjusted for the number of pictures.

Table 2.3: Summary of the monthly maximum implied picture quality

	max quality	stdev	max favs	corr. with qual.	corr. adj. for nr pic
0 year of expr	0.619	1.503	0.636	0.583	0.558
1 year	1.025	1.512	0.749	0.593	0.529
2 years	1.235	1.574	0.819	0.584	0.507
3 years	1.299	1.590	0.867	0.568	0.500
4 years	1.302	1.570	0.878	0.538	0.481
5 years	1.264	1.560	0.882	0.525	0.476

Notes: This table summarizes the individual-monthly maximum of the inferred picture quality (Section 2.3.3), which is treated as data in the subsequent analysis. Monthly maximum refers to quality of the best picture *captured* in the given month, by an individual. Years of experience is defined as number of years from the first in-sample picture to the current month of picture-taking. The first two columns summarize its mean and standard deviation. The third column presents average of the highest rating ("favorites") one gets for pictures taken in the month, given that the highest rating is non-zero (15% of the individual-month data). The fourth column presents its correlation coefficient with the highest inferred quality. Finally, the fifth column presents this correlation adjusted for the number of pictures taken (or has favorites) in the given month.

2.3.4 Camera ownership

I next infer camera ownership from the Exif data behind each picture. As previously mentioned, the camera identity is embedded in the picture's Exif data. I then assume that, for a given individual, whenever one observes a new camera capturing its first in-sample picture, I assume that the previous camera has been replaced.

For 75% of all individual-camera combinations, I never observe an old camera taking pictures after the arrival of a new camera. For the remaining 25%, although the earlier cameras still take pictures, the majority of the pictures are taken by the most recent cameras acquired. Figure 2.15 in the Supporting Material documents this, and shows that there are few cases when the latest camera is not the most active one.

2.3.5 Computing price indices

For digital cameras, as other consumer electronics, prices vary a lot among retailers, fluctuate over short periods of time, and display large differences across first- and second- hand markets. I cannot observe the actual prices that the consumers observe. Instead, I observe the monthly average Ebay auction price for each camera model. This price data are averaged across first and secondary markets. Especially for older camera models, it better represents the prices the consumers face compared to a retailer's list price.

With this data, I first deflate prices to 2005 US dollars. Then, separately for both camera formats, I take the weighted average of the prices of all available cameras in a given month, by their market shares in the Ebay auction data.[18][19] Since the data only ranges from 2006 onward, I interpolate the missing values before 2006, by taking a log-linear fit against time, plus a *simulated* regression error.[20]

2.4 Descriptive analysis of consumer experience, camera technology and their interactions

2.4.1 Overview

This section presents reduced form, descriptive analysis on how picture quality evolves in time, in a consumer's experience in photography, and in camera technology. I first present some crude descriptive evidence, which hints that four key themes that are captured by our structural model. First, picture quality increases with consumer experience, suggesting the role of consumer's human capital in photography. Second, camera technology – reflected in the format (DSLR v.s. compact camera) – contributes to picture quality. Third, human capital complements camera format. Lastly, camera switching induces destruction in part of the human capital.

Given the emphasis of consumer human capital in this paper, we then further study the first and last points. In Section 2.4.3, we show that the evolution of each individual consumer's experience is stochastic, in the sense that learning is generated by persistence in picture quality shocks. This suggests that a learning-by-experimentation model will fit

[18] That is, the number of auctions for a given camera model, as a percentage of the total number of auctions in the sample.

[19] In this version, I do not consider other camera characteristics such as resolution. In robustness-check versions when this was considered, I use the same method to compute resolution indices.

[20] Separately for each format, I regress log price index on a linear time trend, and interpolate the missing value using the linear prediction plus a simulated prediction error. The R-squared for the linear regression are around 0.7 for both formats. Keane and Wolpin (1994) use this method to interpolate missing data in their value function calculations.

the human capital evolution path. In Section 2.4.4, I graphically show short-run dynamics in consumer human capital, around the time of camera switching, as direct evidence of consumer switching cost. Finally, I present some patterns in consumer's purchase decisions for new cameras, that is consistent with a demand model with general and product-specific human capital.

One note on selection before the analysis: we can only observe an individual if she kept taking pictures that she would post later, and we cannot observe either an individual's picture quality or her choice of cameras, if she decides to stop updating her Flickr portfolio. Therefore, it is a concern that individuals with lower ability to take pictures might systematically drop out earlier. To address this, I first screen out individuals who I observe for less than 5 years, and only look at the remainder of the sample in their first 5 years. This ensures that everyone in the sample stays at least the 5-year duration. I use this sub-sample to produce the reduced form evidence.

2.4.2 Variations in picture quality to camera technology and consumer experience

In this section, I study basic patterns of picture quality across different time periods and different camera formats. As elaborated in Section 2.3, I measure picture quality as the percentage difference in views, unexplained by the time window when a picture is displayed on Flickr.com. I also take the maximum of picture quality of each individual in a given month, and I argued that this is a measure less prone to selection.

In Figure 2.2, the four panels present four key patterns of picture quality as a function of camera technology and consumer experience.

First, the upper left panel present the quality of an individual's best picture, captured at different points in time. I normalize time as half-years since the capturing of an individual's first picture, that is ever uploaded onto Flickr. The increasing time trend in an individual's picture quality suggests that she gradually learns about how to take better pictures. In addition, concavity of the trend suggests that the more she already learned, the lower her learning rate – or that there is less to learn about. However, from this graph alone, one cannot rule

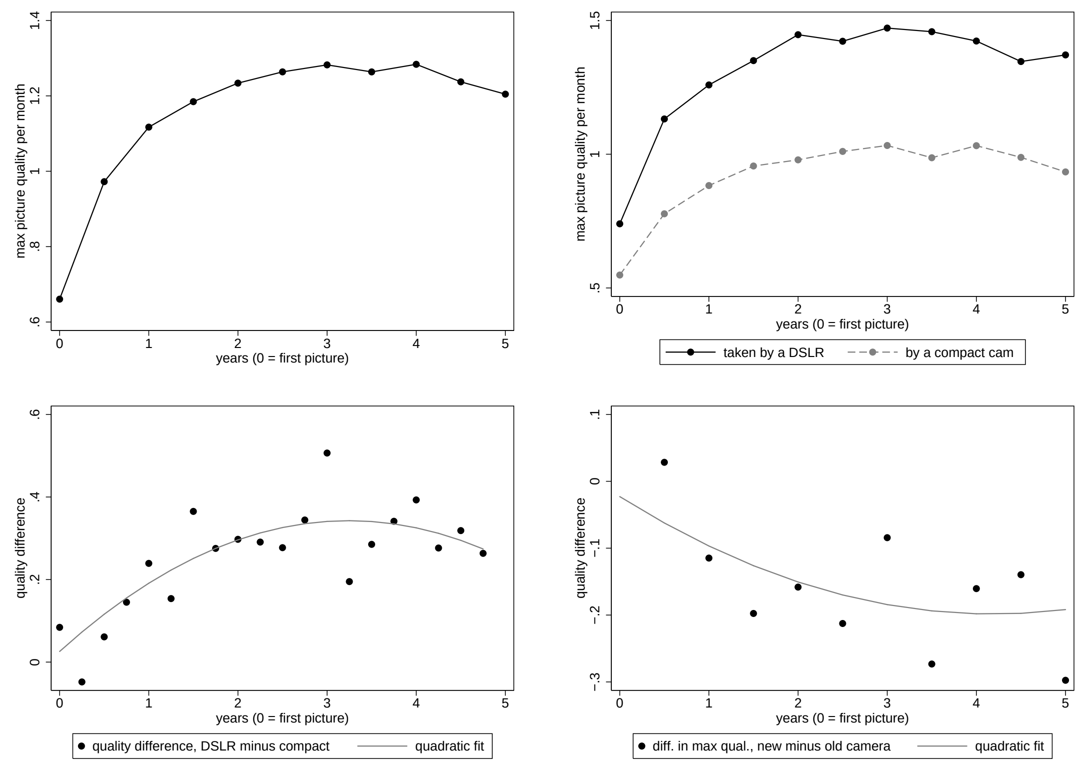


Figure 2.2: Descriptive figures of picture quality on camera and experience

Notes: These four figures summarize some key aspects of picture quality as a function of experience and camera technology. The upper left panel plots the monthly highest picture quality for a given individual, against experience measured in years. The upper right panel plots maximum picture quality conditional on the format of camera taking the picture, without controlling for the (endogenous) choice of camera format. In the lower left panel, I plot *within-individual* differences in their monthly-maximum picture quality, using the two formats of cameras. That is, we focus on a set of individuals who simultaneously use both formats (within a time window of 6 months), and observe the difference in quality. Finally, the lower right panel plots the difference in the picture quality, for an individual using a camera for the first 3 months (which is likely to be a new camera), and the *same* individual a camera for more than 3 months (which is likely to be the previous camera). The difference then represents (negative of) switching cost, and panel shows switching cost as an increasing function of general experience. The horizontal axis, general experience, is defined as the number of years since one's first picture, *ever* posted on Flickr. To eliminate selective attrition, I limit the sample to consumers we observe for more than 5 years. The markers are mean values across individuals, while the solid line is a sample frequency-weighted quadratic fit. Standard errors are not presented, but most visible differences on the figures are statistically significant.

out the possibility that it is the evolution of camera technology that drives up the picture quality.

Next, we condition on the format of camera, i.e. compact camera or DSLR, that takes the picture. The improvement in picture quality over time, conditional on camera format, is presented in the upper-right panel. We find that changes in equipment does capture some of the trend in the previous figure.[21] Conditional on equipment format, the first 3 years in sample seems to improve a consumer's picture quality. Depending on the type of equipment she uses, this quality improvement generates 40-60% more views on her best picture. This implies a large annual growth rate of 12-17%, of a consumer's photography-specific human capital. In addition, in terms of correlation, this also implies that a consumer who chooses a DSLR camera learns faster.

The third panel, on the lower left, extends the previous argument into within-consumer variations. Specifically, I take the difference in quality, between the best pictures per month – taken by the same individual, but – by compact cameras and DSLRs. Because most of the consumers in my data do not use two different camera formats in the same month, I first compute average picture quality by each camera format in a quarter, and then take the difference. We find that having a DSLR camera significantly improves picture quality, but only when the individual has accumulated some experience. In other words, consumer human capital in photography is complementary to camera format. Also, comparing this result to the upper-right panel, we find that the within-consumer difference in quality between the two camera formats is smaller. This implies that consumers who are better in taking pictures choose better cameras, and a cross-sectional comparison of picture quality confounds the selection effect.[22]

Finally, in the above arguments, we have ignored *dynamics* in consumer human capital when switching between cameras. This is addressed in the lower-right panel. Here, I focus on a consumer's picture quality generated by the *same* camera, but contrast pictures taken, in fixed half-year periods, from cameras that the consumer started using for no more than

[21] We implicitly assume that only the camera format is relevant to picture quality. The effect of other camera characteristics will be studied when we discuss modeling of those characteristics in consumer beliefs, in Section 2.5.11.5.

[22] However, the lower-left panel is not clear of all selection biases: in particular, it does not address the endogenous timing in selecting camera formats for a given consumer.

3 months, and cameras that she has used for more than 3 months. Given that we fixed the time period, the camera she just started using are by construction the new cameras, and the others are cameras she used before ("old cameras"). Their difference reflect the negative of consumer switching cost, or loss of camera-specific human capital as reflected in the reduction in picture quality. In addition, I find that the trend of switching cost is decreasing in the duration of experience, suggesting that a proportion of experience, gained from a different camera, cannot be migrated to a new camera.

2.4.3 Stochastic growth in consumer human capital

A deterministic function of human capital on experience seems to capture the increasing and concave learning curve in the aggregate, as shown by the upper-left panel in Figure 2.2. I next test, at the individual level, whether learning by doing can be characterized by such a deterministic experience curve. We find that for a given individual, the majority of an improvement of picture quality comes from carry-over effects of past quality improvements – specifically, those incurred during the previous picture-taking occasion. This result then speaks for a stochastic model for consumer human capital evolution.

In particular, denote a picture taking month m, and I estimate a linear specification of current-period (maximum) picture quality. Picture quality is a function of 1) how many month the consumer spent taking pictures, 2) picture quality in the previous (picture-taking) month, 3) whether the individual uses a DSLR, and 4) calendar time, and allowing for individual fixed effects and a constant term:

$$Q_{ijm} = \theta_0 + \theta_m \cdot m + \theta_q \cdot Q_{ij'm-1} + \theta_t \cdot t_m + \theta_k \cdot SLR_{im} + \theta_i + \vartheta_{im}$$

where t_m denotes calendar time of an individual's m^{th} month of taking pictures.[23] In the above specification, θ_m represents a learning curve, where experience accumulates linearly and deterministically. In contrast to this, θ_q represents a stochastic learning curve, where learning occurs when the previous-period picture quality is high, controlling for hetero-

[23] Implicitly, we assume that the error term is serially uncorrelated, and orthogonal to all right-hand side variables.

Table 2.4: Reduced-form picture quality evolution

	D.quality	D.quality, t-1 (1st stg.)
D.quality, t-1	0.075***	
	(0.006)	
D.dslr	0.183***	-0.070***
	(0.018)	(0.013)
D.time	0.002	-0.024***
	(0.002)	(0.001)
quality, t-2		-0.340***
		(0.006)
constant	0.007*	0.487***
	(0.003)	(0.011)

Note: This table presents Arellano-Bond estimates for Equation (2.3). "D" denotes Δ, or the first difference operator. The constant term represents the experience effect before first-differencing. The endogenous $\Delta Q_{ij'm-1}$ is instrumented by $Q_{ij'm-2}$, and the second column presents first stage estimates. Standard errors are heteroskedasticity robust and clustered by individual. Standard deviation of residuals $\Delta\vartheta_{im}$ is 1.17. This implies that the standard deviation for ϑ_{im} is 0.83.

geneity in θ_i and in the camera format. This is more apparent when we denote Δ as the first difference operator (between m and $m-1$), and rewrite the above into

$$\Delta Q_{ijm} = \theta_m + \theta_q \cdot \Delta Q_{ij'm-1} + \theta_t \cdot \Delta t_m + \theta_k \cdot \Delta SLR_{im} + \Delta\vartheta_{im}. \tag{2.3}$$

We estimate the above specification, by using $Q_{ij'm-2}$ as an instrument for $\Delta Q_{ij'm-1}$. This is a standard way of estimating dynamic linear models.[24] The results are presented in Table 2.4.

I find a positive carry-over from a change of picture quality in the most recent past, to the change of picture quality in the current month. For a 1-standard deviation shock in picture quality in $m-1$, current picture quality increases by 0.06. This is about 1/3 of the average annual growth rate of picture quality, observed in Figure 2.2.

Compared to the stochastic trend, the deterministic trend is rather small and negligible. This finding indicates that a stochastic learning model might be better in characterizing the underlying human capital evolution process. In such a model, new knowledge arrive at

[24]This is because $\Delta Q_{ij'm-1}$ contains $\Delta\vartheta_{im-1}$, which is correlated with the error term $\Delta\vartheta_{im}$ because they have common component ϑ_{im-1}. We can instrument this by $Q_{ij'm-2}$ due to the assumption that ϑ_{im} is serially uncorrelated. See Arellano and Bond (1991) for details.

random quality, and quality-improving knowledge are preserved.[25] The learning by doing model in the next section is a structural representation of this idea.

In addition, camera format effect remains to be large and significant, and the magnitude closely represents the pattern in the descriptive evidence. Finally, time between two picture-taking occasions could reflect forgetting (if it is negative) or learning from other sources (if positive). However, we do not find a statistically significant time effect, and this means that it might not be as important, in this context, to characterize other forms of human capital improvements.

2.4.4 Destruction of human capital when switching across cameras

I now find another way to present evidence for the imperfect transferability of consumer human capital, across different cameras. To do so, I normalize the date of camera-switching to be period 0, and look at (max) picture quality an individual produces in a given month, around the time when she switches between products. Figure 2.3 shows that there is an immediate drop in picture quality at switching. The drop in picture quality indicates that not all the knowledge from the previous product is transferred to the new camera.

In addition, after the camera switch, picture quality quickly goes up in the first 3-4 months, and it *further* gradually increases to a higher level in 1.5 years. This suggests that at the instance of camera switching, the individual loses both explicit knowledge on camera operations (e.g. menu and button layout), as well as implicit knowledge on camera usage (e.g. how to best circumvent a certain product limitation). While the first can be quickly learned in a month or two, the second can only be learned with long experience with the new camera.

Note that the highest picture quality drops despite that the individual produces *more* pictures immediately after switching – as documented by Figure 2.17 in the Supporting Material. If the underlying distribution of picture quality is unchanged, then more pictures should imply a higher draw of maximum picture quality. Also, note that the consumer might

[25]The symmetric structure here also implies quality-destroying knowledge further destroys future human capital. I cannot ensure that this is not a relic of the linear structure (and cannot estimate general a dynamic nonlinear panel data model), and thus do not take this into account in the structural model.

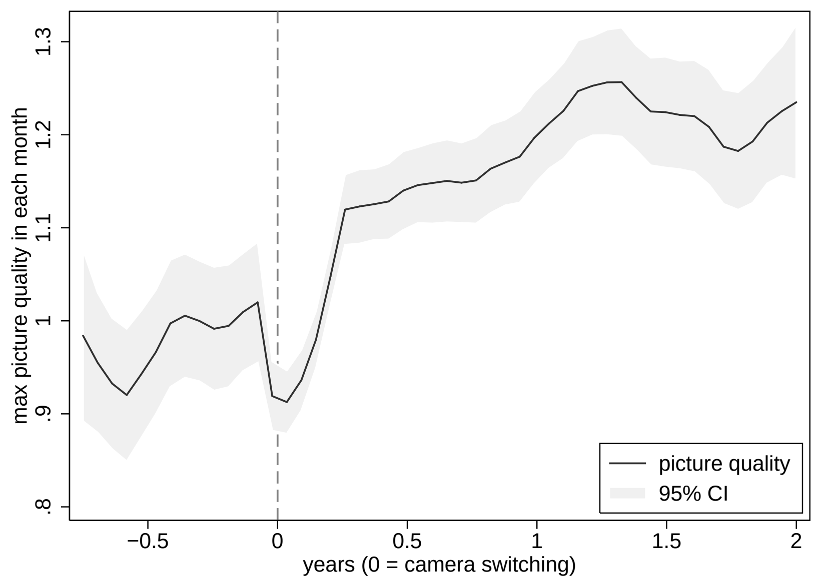


Figure 2.3: Quality of best pictures around camera switching

Notes: This figure depicts the changes in maximum picture quality, around the period when an individual switches cameras. We focus on the years before and after a consumer switches her camera at year 0. With the left vertical axis, the dark line (and shaded areas as its 95% confidence interval) depicts the maximum picture quality that the consumer can produce, using her old camera until year -1/12, and new camera *from year 0*. The line is estimates of local polynomial regression with bandwidth 1.

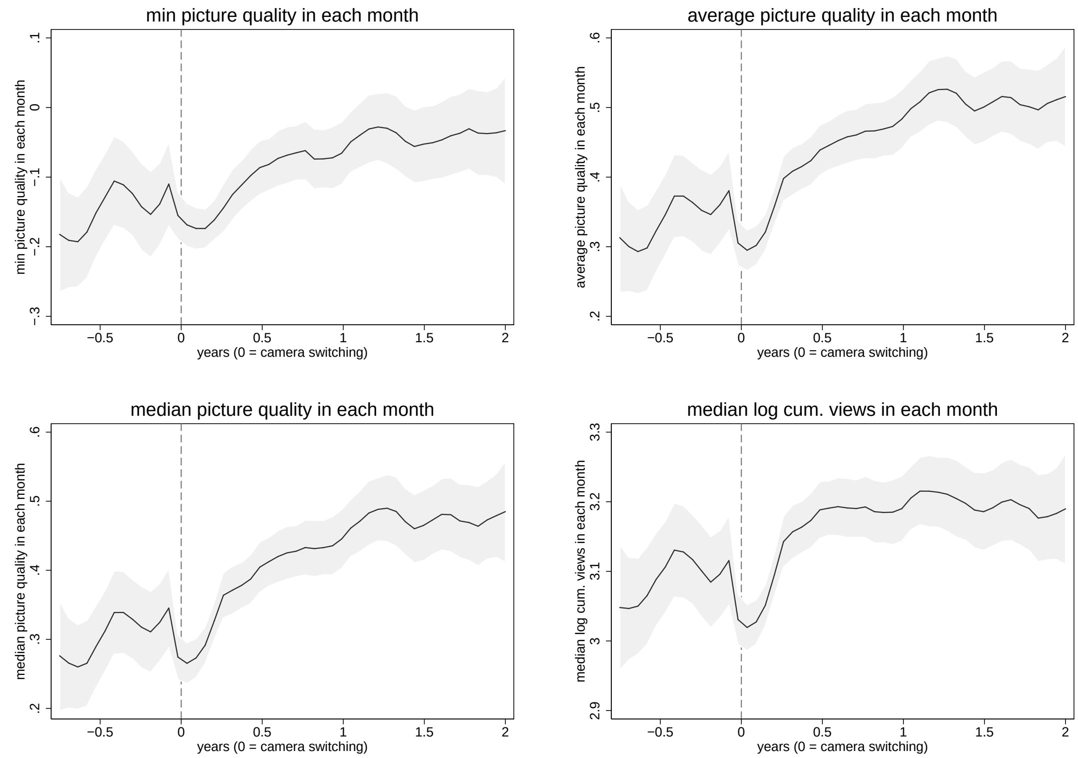


Figure 2.4: Switching cost under alternative measures of quality

Notes: This is a robustness check of changes in alternative measures of picture quality, around camera switching. The first 3 panels are the minimum, mean and median of the picture quality distribution, around camera switching. The lower right figure presents raw data of log views, i.e. not controlling for display time window effects, topic effects and other Flickr-specific effects. We find that the evidence of switching cost is robust to alternative measures.

select away pictures that she considers "bad", especially when she is unfamiliar with the new camera. We estimate switching cost despite potential selection, which might also downward bias the estimate (upward bias the picture quality with new camera). Both concerns imply that this figure presents a conservative estimate of the switching cost.

The pattern is robust (besides being noisier) when conditioning on the direction of camera switching, as presented in Supporting Material Figure 2.16. Also, the pattern is robust when we plot changes in other moments, or order statistics of the picture quality distribution, as well as when we use raw data on views to represent quality. Figure 2.4 presents some robustness checks.

2.4.5 Patterns in consumer choices

Finally, I document some notable patterns in consumer choices of camera format, and their timing of camera switching. These are consistent with a demand system that accounts for endogenous human capital evolution.

I first present evidence that suggests an increasing tendency to use advanced products. To pin down experience effects, I estimate a linear probability model of choice of format, controlling for the role of technology and other sources of *calendar-time* effects, as well as individual fixed effects.[26] The estimates of experience fixed effects are plotted in Figure 2.5. I find that having 5 years of photography experience raises one's tendency to use an advanced camera by 10%. This suggests that experienced consumers are more likely to choose advanced products, given a constant technology level.

Next, I present patterns of consumer camera-switching timing, in Figure 2.6. It is key to control for selective attrition, because otherwise we confound consumers who left the sample with consumers who stop purchasing. By the way I control for consumers who stay for more than 5 years, we ensure that we observe everyone at the end of this sub-sample.

I find that although the probability of camera switching is increasing in the time-in-sample. In addition, and more curiously, even the probability of switching across-brand is increasing in time. We next plot the share of consumers switching across brands, among all switchers, and find a sharply declining profile against time. This is consistent with the previous finding that camera-switching destroys part of the consumer human capital (and brand switching might destroy more), which causes experienced consumers to try to stay with one brand. The increasing profile might be due to evolution of the camera market conditions, such as reductions in prices.

[26] Specifically, I estimate a linear probability model with the choice of camera format on the left-hand side, and experience dummies and calendar time dummies on the right-hand side:

$$Format_{it} = \alpha_i + \sum_{t=1}^{60} \beta_{expr,t} \left(expr_{it} = t\right) + \sum_{y=2001}^{2013} \beta_{year,y} \left(year_{it} = y\right) + \varepsilon_{it}. \tag{2.4}$$

Essentially, controlling for the calendar time effects allows us to compare within a given point in the calendar time, across individuals with different experience stock at this point.

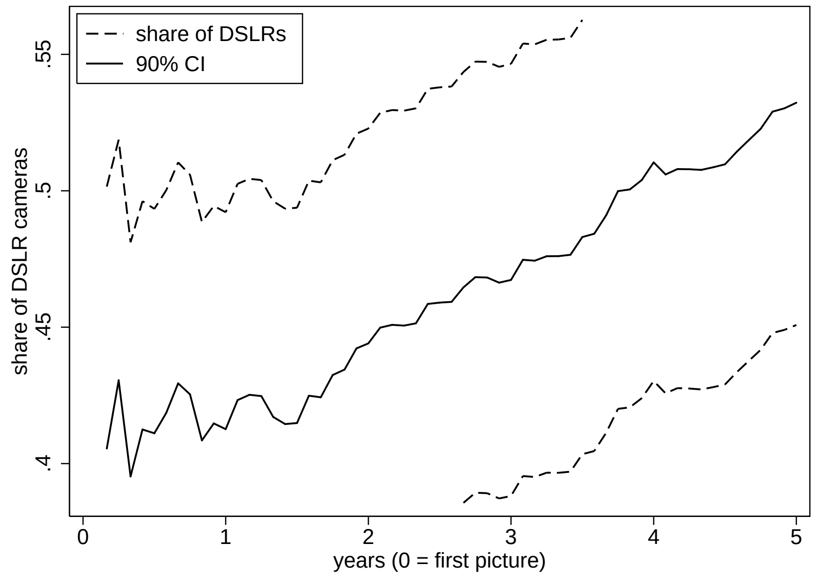
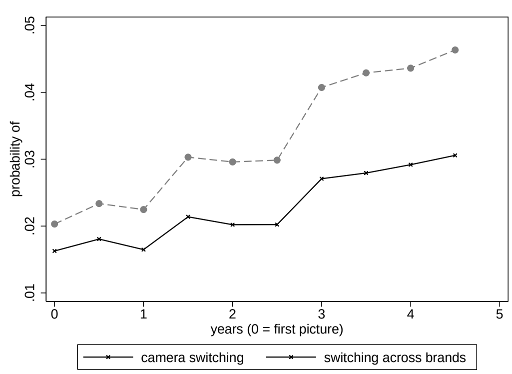


Figure 2.5: Product-format choice and user experience (selected sample)

Notes: This figure depicts the changes in the choice of product format (DSLR vs compact camera), given a user's years of experience – defined as the number of years since one's first in-sample picture. It also depicts (part of) the 90% confidence intervals. To control for selective attrition, I choose the subset of individuals whom I observe for no less than 5 years, and only focus on their first 5 years of data. To control for advances in technology (and other calendar-time effects), I estimate a linear probability model in Equation (2.4), and present the estimated $\hat{\alpha}_i + \hat{\beta}_{expr,t}$ as the calendar-time-detrended estimates in camera usage choice.



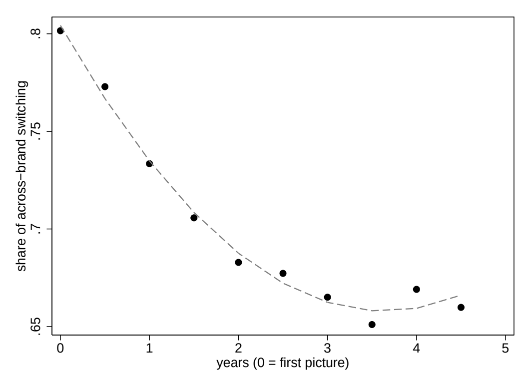


Figure 2.6: Camera-switching and brand-switching probability

Notes: This figure shows the probability of switching between cameras, between cameras of different brands (left panel) and brand-switching conditional on camera switching (right panel). To eliminate selective attrition, I limit the sample to consumers we observe for more than 5 years. So by the end of the figure, the consumers are known to be in the sample. In the right panel, the dashed line is a frequency weighted quadratic fit.

2.5 A structural model

2.5.1 Overview

This section presents the structural empirical model. Whereas the model on durable good purchase with learning by doing is general, it will be presented in the context of digital camera markets for concreteness.

In the model, I jointly characterize a consumer's decisions to purchase digital cameras, and her decisions to use the product. Combining a camera and the stock of experience – or "human capital" – produces pictures that generate consumption utility,[27] and at the same time, contributes to the consumer's human capital stock. Therefore, past usage decisions build up consumer human capital, and hence future utility. With rational expectations and a non-zero discount factor, the consumer makes camera replacement and usage decisions, taking into account the consequences of her decisions on her future human capital stock.

2.5.2 Timing

Consumer i in each period $t = 1, ..., T$ decides whether to purchase a new camera and whether to produce pictures. In the process, her camera quality and human capital – her knowledge of using the camera – endogenously evolve as a consequence of her decisions. The timing of her decisions and state variable evolution is as follows: she first chooses whether to purchase a camera, and in the case of purchase, which *format* and *brand* to buy. Given the brand-format combination, she does not know which camera "model" has higher potential quality, and will randomly draw one according to the market distribution at the time of purchase. If she buys a new camera model, she *immediately* replaces the old one with no resale value. After her decision of camera purchase is made, she decides whether or not to take pictures in this period. If she does so, she experiments on a new method – for example, she tries out a new feature of the given camera – and finds the best way she

[27] I follow the terminology in Michael (1973) and Foster and Rosenzweig (1995). Alternative terminology include "know-how" (Besanko et al., 2010), and "expertise" in Alba and Hutchinson (1987). I also use the term "knowledge" interchangeably with human capital, and this is not to be confused with information.

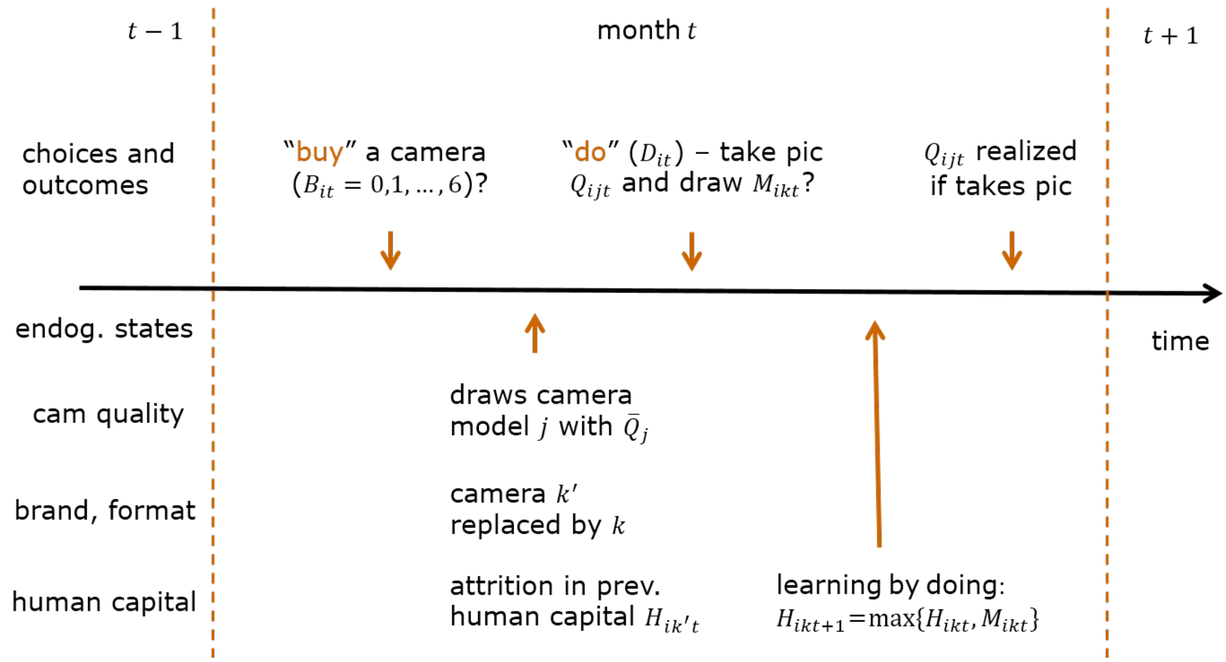


Figure 2.7: Timing of decisions and evolution of the state variables

Notes: The figure presents the timing assumptions of consumer decisions and state evolution, in a given period.

knows to take the picture. Her knowledge on the new method is then incorporated in her experience, potentially improving her human capital. At the end of the period, she derives utility from purchasing a camera, and from the best picture she took, as well as dis-utility from expenditure on the new camera, and effort spent on taking the picture.

I graphically outline the timing assumption in the decision problem, in Figure 2.7. However, many of the notations are introduced throughout this section. Therefore, the figure might be useful for revisiting the timing assumption.

2.5.3 Decisions on camera replacement and usage

I denote *all* consumer decisions in a period as $\mathbf{A}_{it} = (B_{it}, D_{it})$, where symbols A, B, and D stand for "action", "buy" and "do", respectively. To characterize "buy" or B_{it}, because there are over 2000 camera models, I cannot fully realistically characterize the dynamic decision among those within reasonable computation burden. Instead, I characterize purchase decision to be a choice over a brand-format combination, between the two formats (a compact camera or a DSLR), and among three brands (Canon, Nikon and "other brands"). We refer to a brand-format combination as a "camera". The consumer can also choose not to purchase in period t, in which case we denote $B_{it} = 0$.

State variable $K_{it} = 1,...,6$ denotes the camera, owned by consumer i *at the end of* period t.[28] If a camera is purchased, the consumer replaces the previous camera that she owned with the new one, i.e.

$$K_{it} = \begin{cases} K_{it-1} & \text{if } B_{it} = 0 \\ B_{it} & \text{if } B_{it} > 0. \end{cases} \tag{2.5}$$

I do not consider resale, or multiple camera ownership.[29]

Given the decision of purchasing (brand and format) B_{it}, the consumer knows that the camera she will receive has idiosyncratic quality $\bar{Q}_j$, but is not informed of the realization of it before receiving the product. She then expect to draw a realization of $\bar{Q}_j$ fron the market distribution at the time of purchase, and keep it until she purchases another product. As will be discussed in Section 2.5.6, this simplifies the choice problem and avoids modeling of choices over hundreds of camera models.

The binary variable D_{it} ("do") denotes the decision of whether to take pictures ($D_{it} = 1$) or not ($D_{it} = 0$), using the latest camera K_{it} with model j, i.e. after the replacement decision. If she decides to use the camera, she incurs a cost of effort e_i, which summarizes the disutility or utility from taking pictures in a period. Also, she takes *one draw* that determines the realization of the highest picture quality – denoted Q_{ijt} – from which she derives her consumption utility.[30]

To keep the model simple, I do not model the decision on the number of pictures to take. Modeling this aspect will also necessitate modeling of picture selection and upload decisions, which is not identified without stronger assumptions, due to that we only observe the selected pictures. Also, this is not central to the mechanism this paper addresses.

[28] I fix $k = 1,2,3$ to be the compact cameras, and 4,5,6 to be DSLRs. Also, I use $\tilde{k} = 1,4$ as the realized format, where naturally, 1 refers to a compact camera and 4 refers to a DSLR.

[29] As discussed in Section 2.3.4, I observe very few cases of possible multiple camera ownership. This does not justify modeling this, as it will greatly complicate computation. Also, I do not observe resell, or whether a consumer purchases from the first or secondary market.

[30] Because j is sufficient statistics for k, Q_{ijt} is a function of camera k and model j, among other elements.

2.5.4 Production function

The individual derives utility from the quality of the *best* picture she produces, which is an output of three components: first, her personal, time invariant characteristics; second, her (time variant) experience – or human capital – and third, the technology of her camera. I denote her personal characteristics as a parameter q_i, which is time invariant, but can differ across i. I denote her human capital as H_{ikt} (with respect to camera k), the technology of the camera format $\tilde{k}$ as $\gamma_{\tilde{k}}$, and the technology specific to a camera model j as $\bar{Q}_j$.

Combining these components, we specify the production function for the best picture quality of period t, as

$$Q_{ijt}\left(\bar{Q}_j, K_{it}, H_{ikt}\right) = q_i + \bar{Q}_j + \gamma_{\tilde{k}} \cdot H_{ikt} + \eta_{ijt}. \tag{2.6}$$

where η_{ijt} is an independent and identically distributed (IID) error term, that captures non-systematic variation in the maximum picture quality.

Next, I define consumer human capital and its evolution in Section 2.5.5, and specify the evolution of technology $\bar{Q}_j$ in Section 2.5.6.

2.5.5 The evolution of consumer human capital

2.5.5.1 Consumer human capital and learning by doing

Taking pictures requires a camera. However, despite that digital cameras usually have some automatic features, these features ensure the production a picture but does not guarantee its quality. With little or no experience, a consumer can only experiment by randomly choosing a *method*, i.e. a way to take her picture. On the other hand, for a consumer with some experience, her past experience guides her choice of the current best method. Therefore, on average, a consumer with more experience is capable of producing higher picture quality.

To model this, I assume that there is an underlying distribution of methods, normally distributed according to

$$M_{ikt} \sim \mathcal{N}\left(0, \sigma_i^2\right).$$

If a consumer has no experience, she draws M_{ikt} and uses it (and camera k) to produce a picture. She then keeps the method until she finds a better one. The best method she ever found then defines her human capital H_{ikt}.

With some stock of human capital, a consumer who decides to take pictures in the current period will first draw M_{ikt}, and compare this with the best method she knows from the past, H_{ikt}. She then takes the best method between the two, and use it to produce the picture. If the consumer discovers a good method in this period, i.e. $M_{ikt} > H_{ikt}$, not only will the current picture quality take into account the higher draw, but she will remember this higher draw and replace her past knowledge with it. Formally, learning by doing is the replacement of obsolete human capital:

$$H_{ikt+1} = \begin{cases} M_{ikt} & \text{if } D_t = 1 \text{ and } M_{ikt} > H_{ikt} \\ H_{ikt} & \text{otherwise.} \end{cases} \tag{2.7}$$

Note that this equation recursively defines consumer human capital.

In fact, this way of modeling human capital evolution is closely related to Lucas Jr and Moll (2014).[31] This model of human capital and learning by doing generates three attractive implications. First, higher current human capital stock leads to higher *expected* picture quality tomorrow, using the same camera. Second, if the consumer takes pictures every period, the rate of discovery for better methods, $\Pr(M_{ikt} > H_{ikt}|H_{ikt})$, is decreasing in the human capital stock. This implies decreasing learning speed in experience, which is a common feature in many learning models. For example, the Bayesian learning model in Erdem and Keane (1996), the Bayesian learning by doing model in Jovanovic and Nyarko (1996), and the characterization of production experience curve as in Benkard (2000) and Besanko et al. (2010), all share the same feature.[32]

Thirdly, as an important difference, this characterization of learning by doing generates an asymmetric effect of past picture quality *shocks* to current picture quality. In fact, a

[31] In their paper, an individual decides whether to work or learn. In the second case, she takes a draw from the human capital distribution, and adopts the drawn human capital only if it is higher than her own.

[32] In fact, two earlier versions of this paper used, respectively, the Bayesian learning by doing model, and a reduced-form experience curve with decreasing return. The main result stays the same with this version.

positive picture quality shock contains a persistent part, which contributes to future picture quality. On the contrary, a negative picture quality shock is entirely transitory. In Section 2.4, we documented that changes in picture quality mostly come carry-over of from past changes in picture quality, rather than deterministic time trend. Our modeling approach reflects that result.

As normalizations, we impose that the consumer starts with zero human capital:

$$H_{ik1} = 0.$$

This implies that a consumer will always drop all negative draws of method. We also normalize the maximum attainable human capital to 1. This normalization is required because both learning speed, switching cost and the returns to human capital in picture quality are free parameters.

2.5.5.2 Switching cost

We discussed how a consumer learns by using the same camera k. If, however, she decides to switch to camera k from camera k', not all her knowledge from k' is transferable to the new camera. For example, the menu layouts of one camera is different from another, and even if a consumer knows to apply a certain method, the extra time spent figuring out how to change settings might cause her to miss shots.

Formally, there is attrition on a consumer's human capital when she switches from camera k' to k. Note that here, for notation simplicity, we implicitly denoted the *end-of-period* camera brand-format: $K_{it} = k$ and $K_{it-1} = k'$. With this notation, I impose that the switching cost is proportional to the current human capital stock:

$$H_{ikt} - H_{ik't} = -s_{k'k} \cdot H_{ik't}, \tag{2.8}$$

The proportionality structure is motivated by Figure 2.2 (lower-right panel), where, the picture quality *contrast* between a consumer's previous and new camera is increasing in the years since her first picture. This suggests that consumers with a longer history of picture taking have accumulated much experience specific to cameras that they are familiar with. Finally, switching cost $s_{k'k}$ is set to be symmetric in k and k', to reduce the number of parameters to be estimated; and I will further impose some restrictions for the same purpose.

2.5.6 Camera technology

The quality of pictures produced by the individual also depends on the technology of her camera. Specifically, camera technology plays two roles in the production of pictures:

First, the *format* of camera, i.e. compact camera or DSLR, complements a consumer's human capital. As shown in the production function (Equation (2.6)), I model camera format effect to be a parameter on consumer human capital – denoted $\gamma_{\tilde{k}}$ – in her production function.[33] $\gamma_{\tilde{k}}$ is constant across consumers and time, and known to all consumers. This means that with higher human capital, the same improvement in camera technology generates larger changes in picture quality, which is reflected in Figure 2.2.

Second, the characteristics of the camera model j, $\bar{Q}_j$, affects the *level* of picture quality she produces. To the researcher, since I am able to observe the productivity level of all products using this data-set, I model $\bar{Q}_j$ as an index of camera resolution (in integers of mega-pixels), and the year of introduction:

$$\bar{Q}_j = \sum_{r=1}^{35} \psi_{1,r} \mathbf{1}\left(resolution_k = r\right) + \sum_{y=2000}^{2013} \psi_{2,y} \mathbf{1}\left(year_k = y\right). \tag{2.9}$$

I discuss implementation of this in Section 2.5.11.5. On the other hand, the estimates of this is not known to the consumers. Therefore, as $\bar{Q}_j$ is not known before the purchase of j, the consumer forms a belief that $\bar{Q}_j$ is drawn from a market distribution at the time of purchase,

$$\bar{Q}_j \sim \mathcal{N}\left(\bar{q}_t, \sigma_q^2\right). \tag{2.10}$$

That is to say, before purchase, the expected quality that an individual will get depends on the time of purchase, but the exact quality is uncertain. If the consumer does not purchase,

[33]Recall that we denote camera format $\tilde{k} \in \{1,4\}$ to denote, respectively any compact camera and any DSLR.

the past camera j' stays, with quality $\bar{Q}_{j'}$. This then implies that, holding j' fixed, purchase decision will display positive duration dependence, such that a consumer will be more likely to purchase a new camera, the longer she holds her previous camera. Note that this duration dependence is generated without camera quality depreciation.

Similar to Gowrisankaran and Rysman (2012) and Hendel and Nevo (2006), our way of modeling technology index is a dimensionality-reduction assumption. However, different from them,[34] technology in our model is not a Markov process, but rather a times series of normal distributions with constant variance. Intuitively, in the context of digital cameras, a Markov process generates the undesirable prediction that the "technology frontier" might "shrink" if there is an unfavorable draw; on the other hand, a deterministic model can better represent how the frontier grows.

An undesirable feature of this specification, however, is that the distribution of $\bar{Q}_j$ is non-stationary, which will generate a non-stationary choice problem. We address this in Section 2.5.8.4.

2.5.7 State space and flow utility

We first clarify the relevant state variables before presenting the utility specification. At the beginning of period t, the consumer decisions depends on the camera she owns at the end of last period, $K_{it-1} = k'$, and the quality of the model j' owned at the end of last period, $\bar{Q}_{j'}$. Her decision also depends on her human capital stock with respect to camera K_{it-1}, denoted $H_{ik't}$. Finally, there are three exogenous state variables, not presented in Figure 2.7. The first two are two prices, $P_{\tilde{k}t}$ for $\tilde{k} = 1,4$, each specific to a camera-format $\tilde{k}$. The third one is calendar time, which is only relevant in shaping the distribution of $\bar{Q}_j$, as discussed in Section 2.5.6. We denote $\mathbf{S}_{it} = \left(\bar{Q}_j, K_{it-1}, H_{ik't}, \mathbf{P}_t, t\right)$ for compactness of notation.

With the state space clarified, we present flow utility. In the model, the consumer derives per-period utility from from purchasing a camera, and from *consuming* the quality of the best picture she took. Also, she derives dis-utility from the money she spent on the new camera, and from the effort on taking the picture. If she does not take pictures or purchase

[34]Gowrisankaran and Rysman (2012) assume that the discounted sum of future utility is Markov, while Hendel and Nevo (2006) assume that a part of the individual flow utility is Markov.

cameras, she derives utility zero plus a random shock. The flow utility $\tilde{u}_{it}(\mathbf{A}_{it},\mathbf{S}_{it})$, is then constructed based on the four parts, plus utility shock:

$$\begin{aligned} u_i(\mathbf{A}_{it},\mathbf{S}_{it})+\varepsilon_{it}(\mathbf{A}_{it}) &= (\alpha_i\cdot\mathbb{E}[Q_{ijt}|B_{it},D_{it},\mathbf{S}_{it}]-e_i)\cdot\mathbf{1}(D_{it}=1)+ \\ &\sum_{k\neq 0}\left(\beta_1 P_{\tilde{k}t}+\beta_2\left(P_{\tilde{k}t}\right)^2\right)\cdot\mathbf{1}(B_{it}=k)+ \\ &\sum_{k\neq 0}\sum_{k'}\lambda_{i,k'k}\mathbf{1}\left(B_{it}=k,K_{it-1}=k'\right)+\varepsilon_{it}(\mathbf{A}_{it}). \end{aligned} \tag{2.11}$$

In the above specification, the first term characterizes the *expected* utility for producing picture quality Q_{ijt}, without knowing its the exact realization. Although we allow the expected utility as a function of all states and actions, it only depends on the brand-format of the previous camera, $K_{it-1}=k'$, purchase decision B_{it}, human capital of the previous camera $H_{ik't}$ – which, together with B_{it}, implies human capital for the current camera – and camera quality $\bar{Q}_j$, which is a function of past camera quality $\bar{Q}_{j'}$ and time t. I allow the marginal utility on picture quality to be heterogeneous across individuals. Denote it α_i. Also, parameter e_i characterizes the effort cost in the attempt to experiment a method and produce the picture(s).

The second term captures the conventional price effects in the consumer purchase decisions. Specifically, I impose quadratic dis-utility from the price spent, when the individual purchases a new compact camera ($\tilde{k}=1$) or a DSLR ($\tilde{k}=4$). Because the prices of compact cameras are very different from the prices of DSLRs, one can imagine that the marginal dis-utility from spending an extra dollar might be different on a 80-dollar compact camera, and on a 800-dollar DSLR camera. I allow for a quadratic specification to capture the difference in the marginal dis-utility.[35] To ensure that marginal utility does not change sign within the support of observed prices, I impose a restriction that the turning point of the U-shape does not go below 800 dollars.

Finally, the third term in the utility specification characterizes the immediate (dis-)utility in purchasing a new camera. This include, for example, the psychological effect of choosing

[35] A linear specification will not fundamentally change estimates of the other parameters, but will predict very different elasticities for DSLRs and for compact cameras. On the other hand, a natural log specification will overly flatten the dis-utility profile, within common price range for DSLRs.

a brand that is different from the current camera, or a status effect from purchasing a DSLR (regardless of the quality of pictures one can generate), etc. Further restrictions are placed in Section 2.5.11.

2.5.8 Transition probability of the state variables

2.5.8.1 Human capital

As explained in Section 2.5.5, human capital improves when the consumer decides to take pictures in a period, *and* discovers a better method. So if the consumer does not take pictures, human capital stays constant for a period. If she takes pictures, *but* does not discover a better method, human capital also stays equal to the previous period value. Therefore, following Equation (2.7), given picture taking, the conditional probability density function for the next period human capital to be equal to h, is

$$\begin{aligned}\Pr\left(H_{ikt+1}=h|H_{ikt},D_{it}=1,B_{it}=0\right) &= 0\cdot\mathbf{1}\left(h<H_{ikt}\right)+\\ &\quad \Pr\left(M_{it}\leq h\right)\cdot\mathbf{1}\left(h=H_{ikt}\right)+\\ &\quad \phi\left(h/\sigma_i\right)\cdot\mathbf{1}\left(h>H_{ikt}\right),\end{aligned}\tag{2.12}$$

where the first term indicates that human capital in $t+1$ cannot go below H_{ikt} if there were no camera switching; the second term indicates that if the method draw was "unlucky", that the consumer did not find a better method than the historical best, human capital will stay at H_{ikt}. The last term captures the distribution of improvement, where ϕ denotes the standard normal probability density function, and one can rewrite this term into $\phi\left(h/\sigma_m|M_{it}>h\right)\cdot\Pr\left(M_{it}>h\right)$ keeping $h>H_{ikt}$. Here, it is clear that the density of improved human capital depends on whether the consumer could find a better method, and the conditional density of "better methods".

With camera switching from k' to k, human capital first takes a loss due to switching cost, and then undertakes Equation (2.12). That means, if signal exceeds $\left(1-s_{k'k}\right)H_{ik't}-$

which is the "left-over" human capital after switching – learning will happen. This implies

$$\begin{aligned}\Pr\left(H_{ikt+1}=h|H_{ik't},D_t=1,B_{it}=k\right) &= \Pr\left(M_{it}\leq h\right)\cdot\mathbf{1}\left(h=\left(1-s_{k'k}\right)H_{ik't}\right)+\\ &\quad \phi\left(h/\sigma_i\right)\cdot\mathbf{1}\left(h>\left(1-s_{k'k}\right)H_{ik't}\right). \qquad (2.13)\end{aligned}$$

Note that ϕ – the density of new method arrival – is unchanged. This highlights the assumption that human capital does not alter the underlying method distribution. However, the probability that human capital improves has changed, because the area $h>\left(1-s_{k'k}\right)H_{ik't}$ is now larger. Therefore, compare terms between Equations (2.12) and (2.13), although the expected future human capital decreases as a result of switching cost, the expected learning rate increased. In particular, if the switching cost is large, this implication resembles the sharp drop in picture quality after camera switching, but the high learning speed that immediately follows.

2.5.8.2 Camera

Besides human capital, there are two terms that changes with a camera: the brand-format combination K_{it} (which we refer to as a "camera") and the characteristics-specific quality index $\bar{Q}_j$ (which we refer to as a "model"). We keep track of all possible current and future brand-format combinations. Therefore, the camera evolves deterministically, as characterized in Equation (2.5).

The quality index stays constant if the camera, say j', does not change. If the consumer switches to a new camera j, as indicated in Equation (2.9), she draws a new $\bar{Q}_j$ from an exogenously evolving distribution that depends only on calendar time of purchase. This pins down the distribution of $\bar{Q}_j$:

$$\Pr\left(\bar{Q}_j<q|B_{it}\neq 0,t\right)=\Phi\left(\frac{q-\bar{q}_t}{\sigma_q}\right) \qquad (2.14)$$

where $\bar{q}_t$ and σ_q are parameters that are estimated in reduced form, and Φ denotes standard normal CDF.

2.5.8.3 Prices

Price transition matrices are exogenously given, and only depend on the price of the same format of camera in the current period. That is to say, we allow for two price transition matrices $\Pi_{\tilde{k}}$, where each element of $\pi_{\tilde{k},ij}$ is

$$\pi_{\tilde{k},ij} = \Pr\left(P_{\tilde{k}t+1} = p_j | P_{\tilde{k}t} = p_i\right)$$

where p_i, p_j are discrete grid points of price.

2.5.8.4 Calendar time

Time is only relevant in characterizing the evolution in the distribution of $\bar{Q}_j$. A fully rational consumer, who takes into account the evolution of time, will have the incentive to wait. This is because the expected camera quality drawn tomorrow will be higher. However, having time in the state variable then generates a non-stationary choice problem, which cannot be solved by iterating on a stationary Bellman Equation.

To solve this, I notice that the evolution of quality index distribution across months is negligible, and therefore ignore the evolution of t in consumer expectations. In fact, if we measure technology as the percentage contribution in the number of views, then *on average*, a camera adopted 1 month later is capable of generating 0.08% more views.[36] However, compared to the change in the $\bar{q}_t$, cameras adopted in the *same month* have a standard deviation of $\sigma_q = 7\%$ in the unit of views. That means, a "lucky draw" of model that is two standard deviation above the mean can generate 14% more views, compared to the "average" cameras adopted in the same period. With this amount of cross-sectional heterogeneity in $\bar{Q}_j$, a tiny shift in the mean (equal to 1% standard deviation) is negligible. Therefore, I safely assume that the consumer takes the distribution of $\bar{Q}_j$ today when forming the expectation for camera quality tomorrow.

Note that this assumption only takes away time evolution between two neighboring months (which is all we need to have a stationary Bellman Equation), but allow time as

[36] See estimation results in Section 2.6.

a relevant state variable in consumer choice probability. For example, our model captures that a consumer's probability of camera switching depends positively on the age of the current camera.

2.5.9 Dynamic programming

With rational expectations, the individual makes purchase and usage decisions every period by maximizing the sum of discounted flow utilities, or solving

$$\max_{\mathbf{A}_{i\tau}} \sum_{\tau \geq t} \delta^{\tau-t} \mathbb{E}_t \left[u_i \left(\mathbf{A}_{i\tau}, \mathbf{S}_{i\tau} \right) + \varepsilon_{i\tau} \left(\mathbf{A}_{i\tau} \right) \right].$$

Given stationarity assumptions on the function $u_i(\cdot,\cdot)$ (as in (2.11)) and transition process of $\mathbf{S}_{it} = \left(\bar{Q}_j, K_{it-1}, H_{ik't}, \mathbf{P}_t, t \right)$,[37] this is a standard dynamic decision problem in spirit of Rust (1987) and others, where the consumer solves the equivalent static decision problem

$$\max_{\mathbf{A}_{it}} U_i \left(\mathbf{A}_{it}, \mathbf{S}_{it} \right) + \varepsilon_{it} \left(\mathbf{A}_{it} \right)$$

where the choice-specific value function $U_i \left(\mathbf{A}_{it}, \mathbf{S}_{it} \right)$ is defined by the Bellman equation

$$U_i \left(\mathbf{A}, \mathbf{S} \right) = u_i \left(\mathbf{A}, \mathbf{S} \right) + \delta \cdot \mathbb{E} \left[\max_{\mathbf{A}'} U_i \left(\mathbf{A}', \mathbf{S}' \right) | \mathbf{S}, \mathbf{A} \right]; \tag{2.15}$$

and all state transition probabilities, introduced in Section 2.5.8, apply in the expectation operator.

2.5.10 Identification

Section 2.3.3 discussed identification of implied picture quality, from cross-sectional data of picture taking and posting dates, and their cumulative views. The key identifying assumptions imposed there are that upload date is exogenous given picture taking date, and that the accumulation of potential viewer base (those who might decide to see the pictures,

[37] Note that we imposed that the transition of t across two periods can be ignored.

depending on quality, popularity or topics) is determined by calendar time but not individual characteristics. We provide supportive evidence on the first assumption and robustness check for the case when the second assumption fails.

I now discuss parametric identification of the structural model, given (max) picture quality, choices of picture taking and camera purchase, and other observed state variables as data. First, given a correct model for camera and picture taking choices, the production function intercept q_i is identified by the initial period observed picture quality, given the normalization of initial human capital at 0, some normalization of $\bar{Q}_j$ and enough variation among initial camera characteristics. In fact, I pre-estimate Equation (2.16) to obtain the implied $\bar{Q}_j$, so as to guarantee identification of q_i. Next, camera format effect $\gamma_{\tilde{k}}$ are identified by comparing differences in the stationary picture quality, across camera formats, because the cap of human capital is normalized at 1. Then, learning speed is identified by observing changes in picture quality given q_i, $\bar{Q}_j$ and $\gamma_{\tilde{k}}$, before human capital reaches its stationary level. Of course, all of these are conditional on a correctly specified choice model.

For the parameters in the choice model, we first identify parameters in the exogenous state transition matrices, $\Pi_{\tilde{k}}$ as price transition of camera $\tilde{k}$, and parameters that capture market technology index evolution, $\bar{q}_t$ and σ_q. They are identified by the observed prices and technology. We also impose that the discount factor δ is known. It is not identified unless with valid exclusion restrictions (Magnac and Thesmar, 2002). Given these parameters and the production side model, utility parameter α_i is identified by variations in human capital and camera (which changes the expected picture quality), on picture-taking decisions. Effort cost e_i is identified from picture-taking choices when the expected picture quality is zero. Price coefficients are identified by price variations, and other utility coefficients are identified by the "left-over" systematic variations in choices; for example, consumers tend to choose the brand she used before, or they tend to purchase a DSLR even when her human capital does not justify so, and so on.

Finally, identification of finite mixture heterogeneity comes from systematic variations in an individual's choices and picture quality outcome. See Kasahara and Shimotsu (2009) for a formal discussion.

2.5.11 Implementation

2.5.11.1 Sources of heterogeneity

To capture heterogeneity in the preferences and the human capital formation processes, I assume that there exists a finite-mixture of permanent individual heterogeneity. The sources of heterogeneity across individuals could come in three (groups of) parameters. First, there is heterogeneity in an individual's production function intercept, q_i (introduced in Equation (2.6)), which characterizes that individuals could differ systematically in their ability to take good pictures. Given that it is unlikely that individuals are born with different experience in photography, this could refer heterogeneity in their experience in photography, prior to registering their Flickr account. Alternatively, one could characterize this as heterogeneity in the initial human capital. However, this approach will require dynamically allocating grid points in human capital, along different trial parameters in estimation.

Second, there is heterogeneity in the variance of a consumer's newly drawn method, σ_i^2. Since an individual can always discard bad draws of methods, a larger variance implies that she is more likely to draw a good method, which then implies higher learning speed. This is also apparent in Equation (2.12), where lower h/σ_i results in larger density.

Third, there is also heterogeneity in several of the utility parameters, including how much a consumer cares about picture quality (α_i), how much dis-utility would be incurred from picture-taking, if the consumer generates zero picture quality (e_i), and how much additional utility (besides the effect on human capital) would be incurred if the consumer purchases, or switches between cameras ($\lambda_{i,k'k}$). The finite mixture setup allows for arbitrary correlation between the three parameters.

Note that we do not allow for heterogeneity in the switching cost $s_{k'k}$ and price coefficients. In previous versions, we allowed for heterogeneity in these parameters, but in most cases do not find clear difference (across segments) in their estimates.

2.5.11.2 Switching cost

To further parameterize the switching cost $s_{k'k}$, I allow it to vary across the cases when the consumer switches within the same format of products, or across formats, or across brands. I assume that switching across formats incurs no smaller switching cost than within a format; and similarly, switching across brands incurs no smaller cost than within a brand. To impose these assumptions, I specify the following structure for the switching cost across formats *and* across brands:

$$1 - s_{k'k} = \left(1 - s^{baseline}\right) \cdot \left(1 - s^{format}\right) \cdot \left(1 - s^{brand}\right)$$

where s^{format} and s^{brand} symbolize the across-format and across-brand switching cost, taking value 0 when the individual switches within format or brand, respectively.[38]

2.5.11.3 Choice intercepts and other explanations of state dependence

The utility function in (2.11) gives a very general specification of choice state dependence and choice-specific intercepts, that does not depend on the potential picture quality one generates. In implementation, I restrict the utility specification to a more parsimonious structure, which is characterized by 5 parameters:

$$\begin{aligned}\sum_{k',k} \lambda_{i,k'k} &= \lambda_{i,DSLR}\mathbf{1}\left(B_{it} \geq 4\right) + \lambda_{i,Canon}\mathbf{1}\left(B_{it} = 1,4\right) + \lambda_{i,Nikon}\mathbf{1}\left(B_{it} = 2,5\right) \\ &\quad + \lambda_{i,FormatSwitch}\mathbf{1}\left(format_{it} \neq format_{it-1}\right) + \lambda_{i,BrandSwitch}\mathbf{1}\left(brand_{it} \neq brand_{it-1}\right)\end{aligned}$$

where λ_{DSLR} captures the immediate utility of purchasing a DSLR camera (relative to a compact camera),[39] $\lambda_{i,Canon}$ and $\lambda_{i,Nikon}$ capture the immediate utility of purchasing specific

[38] For example, if an individual holds human capital stock of 1 and a Canon compact camera, then, switching to another Canon compact camera costs $1 - \left(1 - s^{baseline}\right)$; switching to a Nikon compact camera costs $1 - \left(1 - s^{baseline}\right) \cdot \left(1 - s^{brand}\right)$; switching to a Canon DSLR costs $1 - \left(1 - s^{baseline}\right) \cdot \left(1 - s^{format}\right)$; and finally, switching to a Nikon DSLR costs $1 - \left(1 - s^{baseline}\right) \cdot \left(1 - s^{format}\right) \cdot \left(1 - s^{brand}\right)$.

[39] I cannot estimate a separate compact camera utility because the two brand coefficients almost capture the entire market, so a $\lambda_{i,Compact}$ and $\lambda_{i,DSLR}$ together will produce close-to-perfect co-linearity with the brand parameters.

brands, while $\lambda_{i,FormatSwitch}$ and $\lambda_{i,BrandSwitch}$ capture format- and brand- switching effects (in additional to the switching cost in human capital).

2.5.11.4 Initial conditions

Heterogeneity in the prior-to-sample experience is characterized by the heterogeneous production function intercept q_i.

Choices of the initial cameras are endogenous to preference, initial quality and learning speed heterogeneity. For example, a consumer with higher learning speed might be more willing to purchase a DSLR camera before period 0. Therefore, her DSLR owned at period 1 is not exogenously given. To endogenize the initial cameras, I compute the stationary distribution of camera formats, conditional on segment-specific model parameters and that human capital and market technology are fixed at their initial values. This is similar to Hendel and Nevo (2006).[40]

2.5.11.5 Camera quality index

I capture heterogeneity across cameras of the same brand and format by a state variable $\bar{Q}_j$, which, defined in (2.9), is a function of observed resolution and year of introduction. This implies that the same camera will always have a fixed $\bar{Q}_j$, regardless of who adopts it and when.

From a researcher's point of view, to estimate parameters $\psi_{1,r}$ and $\psi_{2,y}$, so as to infer $\bar{Q}_j$ from observed camera characteristics, I estimate a flexible reduced form model of picture quality, to capture the contribution of different camera characteristics. Specifically, I

[40] Alternatively, one could model the initial brand-format distributions. I only model the initial camera format distributions because, monthly choice probability being close to zero, the brand-format choice probability matrix is more likely to be singular at some parameter values.

estimate

$$
\begin{aligned}
Q_{ijt} \quad = \quad & \sum_{r=1}^{35} \psi_{1,r} \mathbf{1}\left(resolution_k = r\right) + \sum_{y=2000}^{2013} \psi_{2,y} \mathbf{1}\left(year_k = y\right) + \\
& \sum_{e=1}^{60} \psi_{3,e} \mathbf{1}\left(expr_{it} = e\right) + \sum_{\tau=1}^{120} \psi_{4,\tau} \mathbf{1}\left(tenure_{it} = \tau\right) + \\
& \sum_{s=1}^{20} \psi_{5,s} \mathbf{1}\left(cum.switch_{it} = s\right) + \sum_{t_0} \psi_{6,t_0} \mathbf{1}\left(t = t_0\right) + \tilde{Q}_i + \varpi_{ijt}, \qquad (2.16)
\end{aligned}
$$

by regressing monthly maximum picture quality of individual i using camera model j at time t, against indicator variables of camera resolution and year of introduction, and the cumulative number of months that the individual has taken pictures ($expr_{it}$), the cumulative months since the individual appeared in the data ($tenure_{it}$), the cumulative number of times that the individual switched across cameras ($cum.switch_{it}$), and calendar time and individual fixed effects. I then take projected values of the first two terms to be a proxy of the camera j's contribution to picture quality.

I then simplify the individual's belief about $\bar{Q}_j$, by assuming that $\bar{Q}_j$ is not observed prior to purchase, and the individual only knows that it is drawn from the distribution $\mathscr{N}\left(\bar{q}_t, \sigma_q^2\right)$. To get an estimate of this distribution, I plot $\bar{Q}_j$ against time of adoption, for the newly adopted camera. The evolution of the mean $\bar{Q}_j$ and two standard deviation (note that we did not plot standard error of the mean) confidence interval are plotted in Figure (2.18) in the Supporting Material. I find that the average camera quality increases with the time of adoption, while the cross-sectional variation of it stays roughly constant. This motivates the functional form $\mathscr{N}\left(\bar{q}_t, \sigma_q^2\right)$, where only the mean depends on time. The small time trend also motivates why it makes little difference when the individual ignores the evolution of technology in two adjacent months, as discussed in Section 2.5.8.4. I then regress $\bar{Q}_j$ among cameras adopted in month t, against a constant term and linear time trend t. The linear prediction is then taken as the mean:

$$
\bar{q}_t \equiv \hat{\mathbb{E}}\left[\bar{Q}_j | t\right] = \hat{\chi}_0 + \hat{\chi}_1 \cdot t
$$

and the variance of residuals from the above specification is taken as the variance:

$$\sigma_q^2 \equiv \mathbb{E}\left[\left(\bar{Q}_j - \hat{\mathbb{E}}\left[\bar{Q}_j|t\right]\right)^2\right].$$

In structural estimation, these two are treated as known parameters for both the researcher and the consumers.

2.5.11.6 Discount factor

Finally, I give all consumers a discount factor of 0.95 monthly. The discount factor implies that the consumers will discount *away* 70% of the value of a camera in two years, which I find intuitive. This also implies an annual discount factor of 0.54, which is lower than the field-data estimates by Dubé, Hitsch and Jindal (2009) (0.7), but higher than the estimates for the non-durable goods case in Yao et al. (2012).[41]

2.5.11.7 Interpolation of the value function

I discretize several of the continuous state variables into a few discrete values. Because the consumer optimization problem is defined on a high-dimensional state space, curse of dimensionality heavily restricts the number of grid points for each state variables. For example, while the actual price for DSLRs is continuous, I discretize it into a 5-element-set $\{0, 300, 600, 900, 1200\}$.

It is desirable to approximate the value function outside of the discretized state space. In that way, we approximate consumer choice rules at the *actual* state variable, rather than forcing it to the nearest grid point. I apply a simple linear interpolation rule, so that the value function is the weighted average of its values at the two nearest grid points. Take the price of DSLR as an example: value function at the actual price p is the weighted average of its values at p_0 and p_1:

$$V(p) = V(p_0) + \frac{p - p_1}{p_1 - p_0} \cdot (V(p_1) - V(p_0)).$$

[41] The discount factor inYao et al. (2012) is close to zero annually, but is reasonable in their context of mobile phone contracts, since it requires a much shorter-term thinking on the consumer side.

Table 2.5: Estimate of camera quality index transition

	par. est.	std. err.
constant (χ_0)	-0.134	0.002
years since 2000 (χ_1)	0.010	0.000
std dev of residuals (σ_q)	0.072	

Note: Estimates of camera quality transition process, outlined in Section 2.5.6. I first estimate a reduced form of camera contribution in picture quality, as in Equation (2.16). I then take the predicted index of resolution and year-of-introduction as camera quality $\bar{Q}_j$. Next, I find all $\bar{Q}_j$ at the time of camera switching, and estimate its linear specification on the year of purchase. This gives coefficients χ_1 and χ_0. Finally, I predict residuals and compute the standard deviation of the error term.

I interpolate value function on a 3-dimensional state space – $(H_{ik't}, \mathbf{P}_t)$, where the prices are two-dimensional – by nesting three 1-dimensional linear interpolation algorithms. That is, I first interpolate human capital giving other state variables fixed at each other's grid point, then interpolate price of compact cameras fixing human capital at the interpolated values, and all others at their grid points, and then repeat this for the DSLR prices.[42] I find that, even locally, the value function at reasonable parameters is not "smooth enough" to be fitted by hyper-planes. Therefore, interpolating one state at a time (although much slower) is a more precise method than using (local) hyper-planes or (global) multi-dimensional polynomials. The latter follows Keane and Wolpin (1994).

2.6 Estimation results

2.6.1 Transition of exogenous state variables

I first estimate how the state variables transition across periods. In structural estimation, these parameters are known to the consumer and the researcher.

First, I present estimates for the "market" camera quality transition. Implementation of this is documented in Section 2.5.6. We find that the distribution of camera quality index does not vary much in adjacent time periods, but the within-period heterogeneity is large. This motivates the modeling approach of a consumer's expectation of $\bar{Q}_j$ tomorrow.

[42] I do not interpolate technology and calendar year, because these two dimensions are not central to the problem, and the effect of technology ($\bar{Q}_j$) was found small.

Table 2.6: Estimate of transition probabilities for DSLR prices

	"tmr: 0"	300	600	900	1200
"now: 0"	0.999	0.001	0.000	0.000	0.000
300	0.006	0.993	0.001	0.000	0.000
600	0.000	0.013	0.985	0.002	0.000
900	0.000	0.000	0.020	0.976	0.004
1200	0.000	0.000	0.000	0.054	0.946

Note: Estimates of the transition matrix for DSLR prices. This is done before structural estimation, and this matrix is known to all consumers in the model.

Next, I present non-parametric estimates for the price transition matrices. I discretize prices of DSLRs into grids of \$300, and those of compact cameras into grids of \$150. Despite being a rather crude discretization, our way of interpolating the value function outside of state space improves precision of the value functions. We find that prices can only stay the same, or move to the grid point immediately next to the current-period price. In addition, at any price level above zero, there is a higher probability for the price to go down one grid, compared to the probability of going up. Together with consumer's rational expectation of this, our model captures a consumer's incentive to wait for price drops.

2.6.2 Structural parameters

2.6.2.1 Camera format coefficients

Table 2.7 presents structural parameter estimates, for the parameters that are common across segments. I find considerable improvement in picture quality, if a consumer uses a DSLR rather than compact camera. The difference in the two parameters shows that an advanced camera is a strong complement to consumer human capital (and vice versa): for consumers who who reached stationary human capital level, their pictures produced by a DSLR will attract 15% more views, all else equal. This is lower than the difference in picture quality between a DSLR and a compact camera, observed in the descriptive evidence, and it reflects the endogenous choices (and the timing of these choices) of camera format.

Table 2.7: Estimates of common parameters

	parameter	s.e.
Return to human capital - compact camera (γ_1)	0.68	0.03
- DSLR camera (γ_4)	0.83	0.03
Scale of quality error term (σ_v)	0.68	0.01
Switching cost - baseline ($s^{baseline}$)	0.07	0.02
- additional from across formats (s^{format})	0.10	0.02
- additional from across brands (s^{brand})	0.11	0.02
Price/100 (β_1)	-2.04	0.04
Price/100 squared (β_2)	0.13	0.01
Share of low-starter	0.64	0.02

Note: This table reports structural estimates for the parameters that are common across individuals. This includes the share of the first segment. Bootstrap standard errors are reported, which are computed from estimates of 20 random samples with replacement.

2.6.2.2 Switching cost

It is natural to expect that previous knowledge on a specific camera cannot be fully applied to the next camera. The switching cost estimates from 2.8 confirm this guess. I find that switching to a camera of the same brand or format incurs some cost beyond the price, and in addition, any across-brand or across-format switch will incur significantly higher loss in human capital.

For example, for a consumer currently using a Canon compact camera, switching to another Canon compact camera costs 7% of her human capital, while switching to a Nikon compact camera costs 16%. If she decides to upgrade to a DSLR camera, switching to a Canon DSLR costs 17% of her human capital, and 26% for switching to a Nikon DSLR. Since the handling and operation for products of different formats and brands are very different, it is natural to imagine that across-brand and across-format switching is associated with larger human capital attrition. All implied switching costs are statistically significant.

2.6.2.3 Picture quality intercept, learning speed, and the utility for picture quality

We then turn to structural parameters that are heterogeneous across individuals, which are all summarized in Table 2.8. To characterize heterogeneity in learning by doing and the associated camera choice, there are three important parameters: the intercept in consumer

Table 2.8: Estimates of heterogeneous parameters

	"low-starters"	s.e.	"high-starters"	s.e.
Picture quality intercept (q_i)	-0.33	0.04	1.24	0.07
std dev of new methods (σ_i)	0.46	0.02	0.85	0.08
Utility: pref to quality (α_i)	1.77	0.15	1.01	0.05
effort cost (e_i)	0.26	0.08	0.40	0.13
- preference to DSLR ($\lambda_{i,DSLR}$)	4.59	0.23	5.07	0.24
- preference to Canon ($\lambda_{i,Canon}$)	-0.15	0.14	-0.83	0.11
- preference to Nikon ($\lambda_{i,Nikon}$)	-0.50	0.06	-1.31	0.08
- switching formats ($\lambda_{i,FormatSwitch}$)	0.26	0.07	-0.35	0.09
- switching brands ($\lambda_{i,BrandSwitch}$)	-0.22	0.08	-1.40	0.08

Note: This table reports structural estimates parameters that are heterogeneity across-segments. Bootstrap standard errors are reported, which are computed from estimates of 20 random samples with replacement.

production function (q_i), standard deviation of the new method distribution (σ_i), which characterizes learning speed, and the marginal utility for each unit of picture quality (α_i). There are other heterogeneous utility parameters, but they are less central to our analysis.

The constant term in the production function, q_i, is very different across the two segments. Specifically, it is negative for the first segment, and large and positive for the second. Note that I normalized initial human capital to be zero, which is also its lower bound. That means, q_i is the average picture quality in the initial period, net of the camera characteristics index $\bar{Q}_j$ (which is much smaller in magnitude). Hence, I interpret heterogeneity in q_i as differences in an individual's ability to produce picture quality at the start of the sample, and accordingly, I label the two segments "low-starters" and "high-starters", respectively.

Next, the standard deviation of new method distribution, σ_i, captures learning speed. To see this, recall that an individual can always discard bad draws of methods. Hence, a larger variance implies that she is more likely to draw a better method (a higher probability of improving her picture quality each time she takes pictures), as well as larger steps of improvement on average. We find that the high-starters – those with high initial quality intercept estimates – improve much faster and in much larger steps, than the other consumers.

From the estimates of σ_i, we can construct the implied arrival probability of new methods, which captures the rate of, and the expected gain from, learning by doing. Figure 2.8 presents the probability mass of new methods arrival, at each grid point of human capital. From the figure, it is apparent that the low-starters, who have higher variance of new

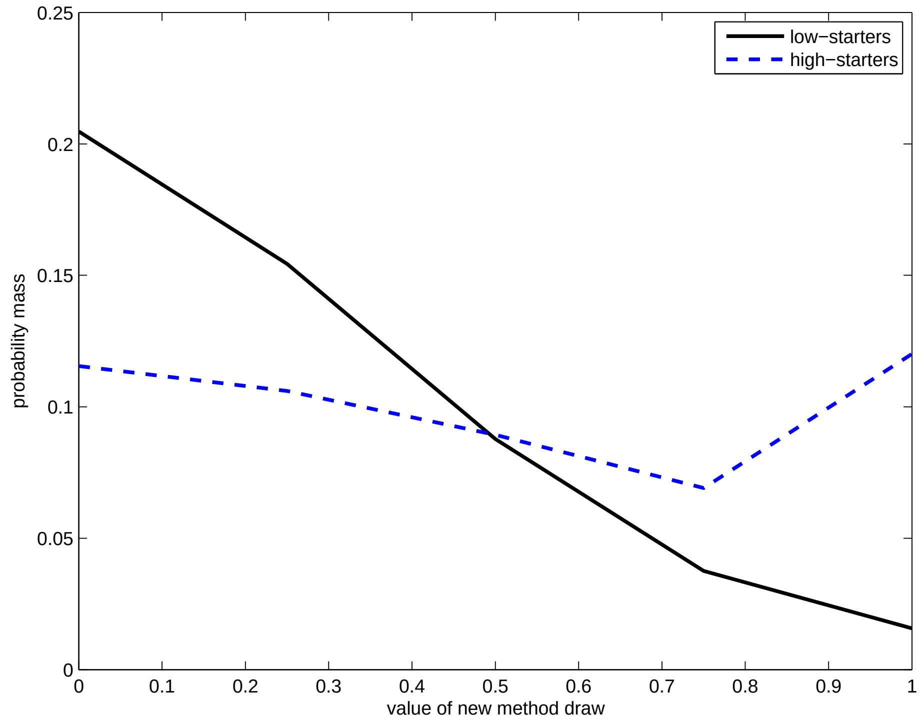


Figure 2.8: Probability distribution of new methods for each segment

Note: This figure shows the probability mass of new methods arrival, at each grid point of human capital, implied by the parameter estimates. For example, if a high-starter is at zero human capital, there is a 0.1 probability that she draws a new method of 0.5, which will become her next-period human capital. Because negative methods do occur and they will be discarded immediately, the area beneath each probability mass function is 0.5. Note that human capital is constrained in $[0,1]$, and draws are larger than 1 are capped at 1.

methods, are much more likely to take large leaps in learning. In addition, the learning rate (i.e. probability of any human capital improvement, upon picture taking) for a high-starter is higher than that of a low-starter. This does not comply with such a hypothesis, that the inherent learning rate for different consumers are similar, and the heterogeneity in q_i reflects differences in initial experience. Rather, this evidence speaks for heterogeneity in the inherent learning rate; that is, the estimate of σ_i truly reflects differences in learning rate, rather than differences of a consumer's location on the same learning curve. Consequently, human capital for the high-starters, who are also fast-learners, converge to their steady state much faster. This is shown in Figure 2.9, where we plot the predicted distribution of human capital by segment.

Finally, the estimates on the marginal utility to picture quality, α_i, also display important heterogeneity across the two segments. We find that the low-starters care much more about

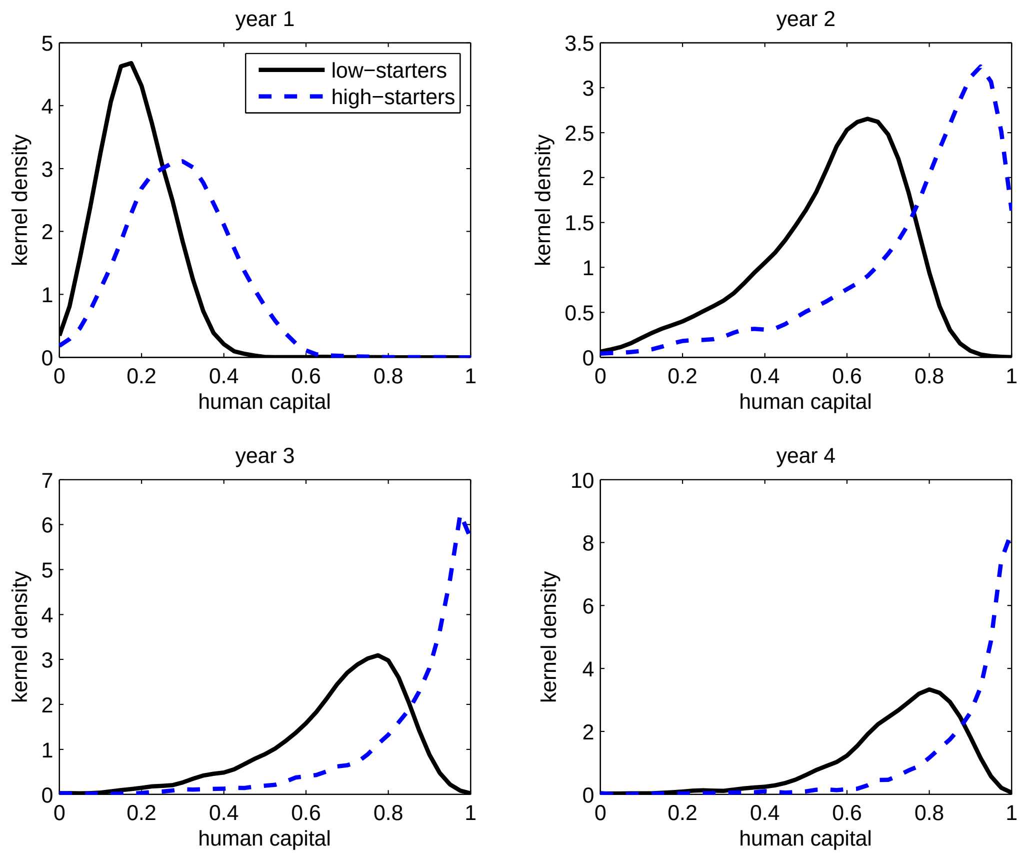


Figure 2.9: Predicted evolution of human capital

Notes: The four panels present four cross-sections of predicted human capital distributions, by segment. Low starters learn much faster, therefore are ahead in their human capital stock. On the other hand, high starters learn slower but has a larger picture quality intercept.

the same unit of picture quality – that is, picture quality that causes the same increase in views – than the high-starters. Hence, although we found that the low-starters are the ones who start low and improve slowly, the same unit of human capital improvement will be converted into camera upgrade – more so than the high starters. In fact, the key source of identification for across-consumer heterogeneity comes from the correlation between a consumer's rate of picture-quality improvements, and her changes in choice probabilities for advanced cameras, over time.

2.6.2.4 Price coefficients

The nonlinear price effects show that, for a one-dollar price change, the individuals are much more sensitive at the lower price range. A consumer becomes insensitive to price changes at 800 dollars, which is over the 95th percentile of the observed DSLR price distribution. In earlier versions, I allowed for different price coefficients for the two segments, and did not find notable differences in their estimates.

2.6.2.5 Other utility parameters

The dis-utility from taking pictures, called "effort", is also different across segments. Intuitively, this parameter rationalizes that a consumer does not take pictures every period, and the difference in this parameter between segments capture that high-starters do not take pictures as frequently as they "should", since for them, taking pictures is much bigger an investment given the higher learning rate. Another way to understand this parameter is to think of it as the cost of obtaining human capital, for each segment. Intuitively, the high-starters can take much more favorable "lotteries" of new methods, but their "price" for the lottery is more expensive, in terms of higher effort cost.

The instantaneous utility parameters from camera purchase and brand switching – that are unrelated to picture quality – show that there is considerably positive utility from purchasing a DSLR camera. This might represent the utility from using the advanced camera features from these cameras, or simply from status effects of using the DSLRs, and rationalizes the tendency to upgrade despite at a low human capital level. Finally, the utility

Table 2.9: Average short-run elasticities

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-2.63	0.03	0.03	0.03	0.03
(2) Nikon Compact	0.01	-2.65	0.01	0.01	0.01
(4) Canon DSLR	0.04	0.04	-3.51	0.04	0.04
(5) Nikon DSLR	0.02	0.02	0.02	-3.53	0.02

Note: This table reports short-run elasticities. I compute elasticities by first calculating the implied choice probabilities for each type of consumer, and then the counterfactual choice probabilities when prices for a given brand-format *in a row* are *temporarily* reduced by 10% for the given month. Then, elasticities are computed from the *averaged* choice probabilities. For example, the first row, second column reads: a 10% temporary decrease in the price of Canon compact cameras *decreases* the demand for Nikon compact camera by 0.3%.

parameters on brand-switching and format-switching are conventionally negative, and captures alternative explanations to state dependence that are unrelated to learning by doing.

2.6.3 Implied price elasticities

2.6.3.1 Short run price elasticities

To verify whether the model produces conventional price effects, I simulate price elasticities from an *instantaneous* 10% price decrease *for a given brand-format*. The change in price is not expected beforehand, and will not persist beyond one month – hence the term "short-run price elasticitities". I calculate the implied choice probabilities for each segment of consumers, with or without the price change, and compute weighted average the choice probabilities, and the implied demand. The price elasticities are then computed from the percentage changes in demand, as responses to the 10% decrease in the prices.

Shown in Table 2.9, I find that the short-run price elasticities are conventional, as in other empirical demand estimation literature in the digital camera industry (Song and Chintagunta, 2003; Gowrisankaran and Rysman, 2012). For example, a 10% decrease in the prices for Canon DSLRs increases the product's current-period demand by 45%. I show the full elasticity matrix in Supporting Material Table 2.15, and it shows that most of the additional demand comes from the consumers who would otherwise not purchase in this period (the "no purchase" category).[43]

[43] On average, the "no purchase" alternative has a baseline market share of 95%.

2.6.3.2 Long run price elasticities

In Table 2.10, I simulate price elasticities from *permanent* 10% price decrease – that is known to all the consumers. The only difference from the short-run elasticities is that price changes are permanent and consumers' rational expectations take this into account.

Table 2.10: Average long-run elasticities

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-2.88	-0.06	0.03	0.08	0.03
(2) Nikon Compact	-0.04	-2.86	0.03	-0.03	0.01
(4) Canon DSLR	0.82	-0.38	-2.77	-0.75	0.04
(5) Nikon DSLR	-0.09	0.94	-0.36	-2.73	0.02

Note: This table reports long-run elasticities, computed from demand responses to permanent price changes of 10% for a given brand-format of camera. For example, the first row, second column reads: a foreseeable, permanent 10% price decrease for Canon compact cameras *increases* the demand for Nikon compact cameras, by 0.6%.

Compare Table 2.10 and 2.9. We find that the effect of a one-time discount is very different from that of a permanent price discount. There are two explanations that jointly determine the pattern we see.

First, because a camera lasts beyond one month, current purchase decisions and purchases tomorrow are inter-temporal substitutes. Therefore, if the price discount is permanent (e.g. for Canon DSLRs), consumers who were thinking about buy other products today, might change their mind and purchase Canon DSLRs in future periods. This is especially true for those who were thinking about buying a compact camera now: when realizing that DSLR prices are cheaper in the future, they might wait for their human capital to reach a certain threshold, and purchase a DSLR only then. In other words, some of the demand for other cameras now could be substituted into the outside option.

We find that that the own-price elasticities for DSLR cameras are higher when a price discount is temporary. That implies, when the individual realizes that the discount is only a one-time offer, she is more likely to take the opportunity right now. This inter-temporal substitution pattern is conventional in dynamic demand models, even in other contexts. For example, in Hendel and Nevo (2006), purchasing now decreases the likelihood of stock-out, and increases future inventory cost, therefore lowering the likelihood of purchasing tomorrow. However, this does not stand for compact cameras.

Second, and less conventionally, we find inter-temporal *complementarity* effect on top of the substitution effect. For example, if the Canon DSLR prices are cheaper in all future periods, investing in compact camera-specific human capital is less attractive. Conversely, investing in DSLR-specific human capital is more attractive. For this reason, a permanent Canon DSLR price drop will further discourage demand in any compact cameras, and encourage purchase of even Nikon and other brand DSLRs.

To see this effect clearly, we show down any direct utility from purchase: that is, additional utility or switching costs from purchase, other than from picture quality that the camera would generate. In other words, we assume that utility only depends on picture quality (which follows the same evolution path) and expenditure. The short-run and long-run elasticities are reported in Table 2.11 and 2.12.

Table 2.11: Short-run elasticities without direct utility from purchase

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-2.61	0.04	0.04	0.04	0.04
(2) Nikon Compact	0.03	-2.63	0.03	0.03	0.03
(4) Canon DSLR	0.00	0.00	-3.57	0.00	0.00
(5) Nikon DSLR	0.00	0.00	0.00	-3.57	0.00

Note: This table reports short-run elasticities, assuming that utility only depends on picture quality and expenditure.

Table 2.12: Long-run elasticities without direct utility from purchase

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-3.27	-0.19	0.12	0.21	0.04
(2) Nikon Compact	-0.26	-3.40	0.11	-0.02	0.04
(4) Canon DSLR	0.01	-0.00	-3.55	-0.01	0.00
(5) Nikon DSLR	-0.00	0.02	-0.00	-3.54	0.00

Note: This table reports long-run elasticities, assuming that utility only depends on picture quality and expenditure.

We find that in this case, for compact cameras, own-price elasticities for a permanent price discount is larger in magnitude, than the elasticities to a one-time price change. For example, consumers purchase even more Canon compact cameras today, when they realize that those cameras are also cheaper in the future. This is to say, current and future purchases of the same camera are inter-temporal complements. We do not find this effect for DSLRs.

In addition, we also find that current purchases of the same format of cameras are also complements, because the consumer foresees that human capital from a Nikon DSLR can easily carry over to Canons, who she expect to purchase more because of the reduced price. Finally, different formats of cameras are now strong substitutes. This is because investments in compact camera-specific human capital cannot be carried over to DSLRs as easily.

There are two side remarks. First, elasticities are higher when we further set the switching costs in human capital to be zero. This implies that imperfect transferability of human capital makes products more differentiated. Second, note that the magnitude of elasticities in Table 2.11 are very different than the benchmark elasticities. This is because we set some utility coefficients to zero, and this drastically changed the choice probabilities.

2.6.4 Model fit

We also examine model fit. In particular, we check whether the model can simultaneously fit the following three patterns. First, we examine evolution in a consumer's picture quality when she slowly gain experience. Second, we check whether the model-predicted shock to her picture quality, at the instance of camera switching, matches the observed patterns in Figure 2.3. Finally, we look at whether a consumer in the model chooses her cameras in the same pattern as the data shows. These are presented graphically in Figure 2.10 and 2.11.

To compute the model prediction of a variable, say picture quality, I first simulate its realizations across all consumers in all periods, for both segments. Then, given a period (a calendar month, or a month relative to camera switching), I average the individual predictions by segment probability and sample frequency.

These figures show that the model fits very well, both for the general trend in picture quality, and for shocks to picture quality around camera switching. This indicates that the learning by doing model can capture both long run and short run evolution of human capital.

Given the characterization of human capital evolution, I find that the model predicts choice probabilities well in general, for purchase of both types of cameras and for picture-taking using the previous camera. The only exception is that it over-predicts the demand for compact camera at the very beginning. This might be due to that consumer human capital

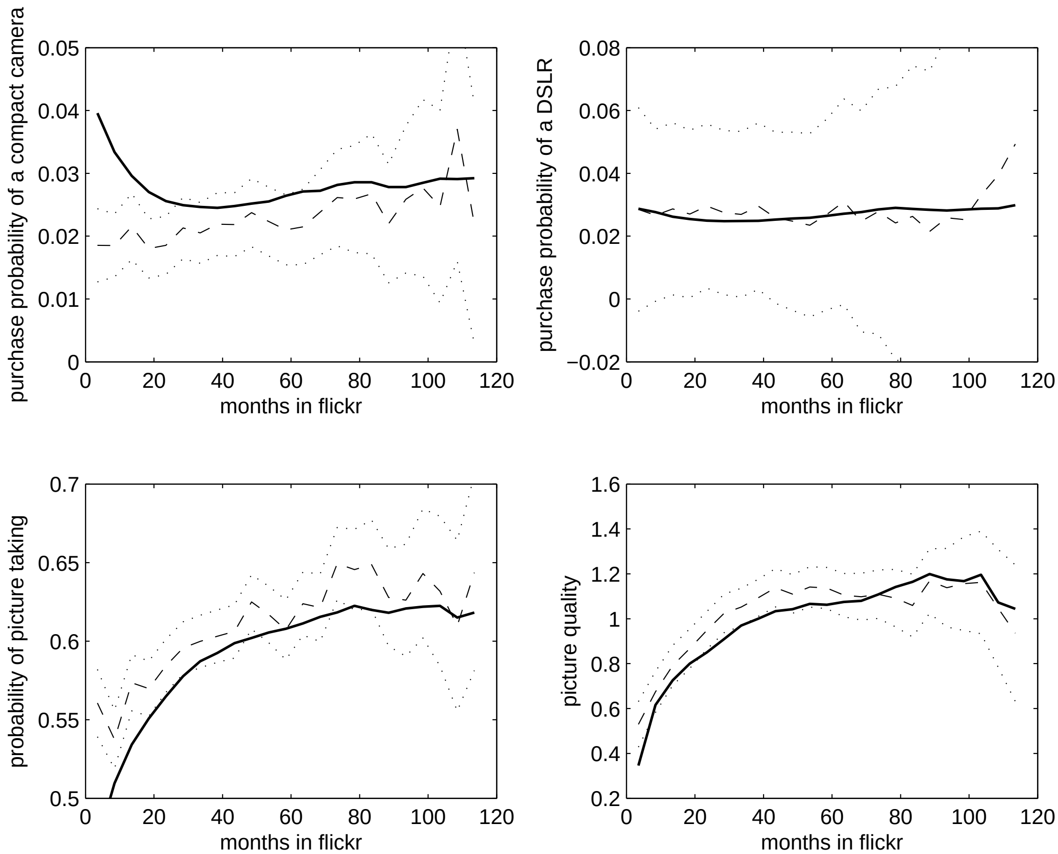


Figure 2.10: Observed and predicted data

Notes: The four panels present model predicted and observed data along a consumer's duration in Flickr. These four panels are, respectively, observed and predicted choice probability of compact cameras and DSLRs (upper panels), observed and predicted choice probability for picture taking, and the observed and predicted maximum picture quality in a month. The predicted values are calculated first by segment, and then averaged across segments by their posterior probability.

is still low, and the model predicts that initial holders of an DSLR camera should switch back.[44]

[44]We endogenized both initial picture quality and initial camera format choice, but we imposed that initial camera is drawn from a stationary distribution given constant human capital at 0. An alternative way of modeling the initial condition might not generate this pattern, but given that this mis-prediction is not severe, we choose not to complicate the model.

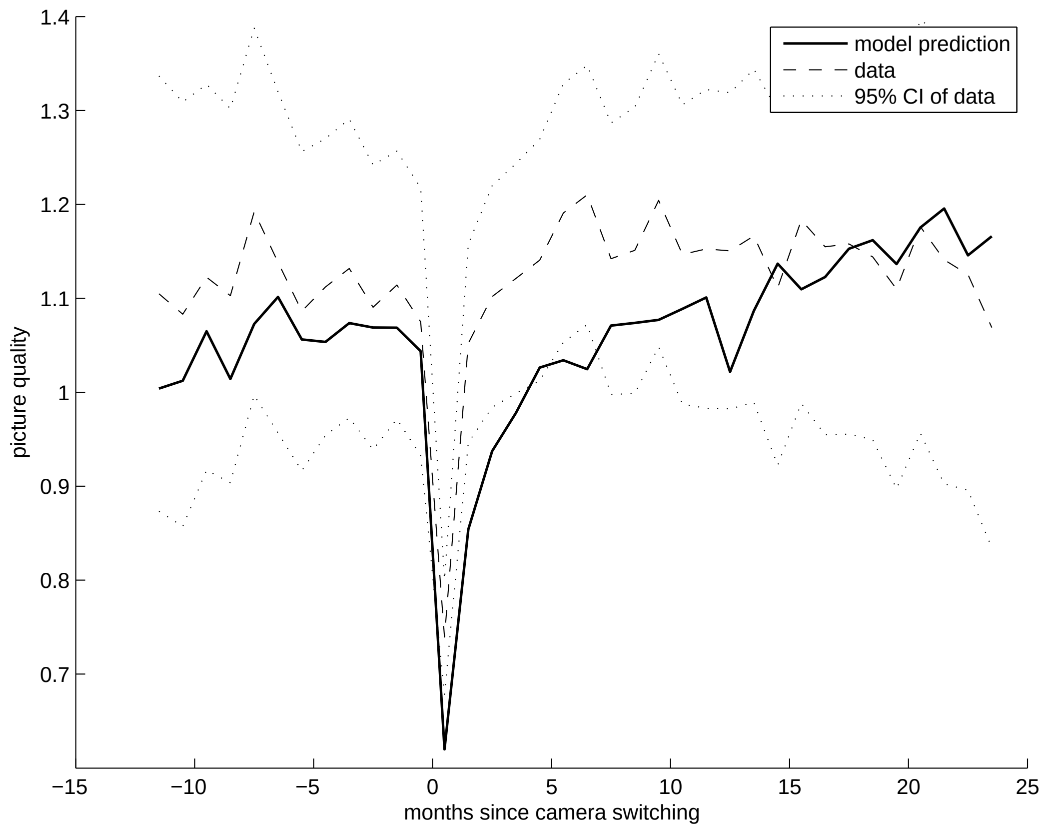


Figure 2.11: Observed and predicted picture quality around camera switching

Notes: This figure presents the observed and predicted maximum picture quality in a month, by months since camera switching. The month that the individual start using a new camera is normalized to month 1. The predicted values are calculated first by segment, and then averaged across segments by their posterior probability.

2.7 Counterfactual implications

2.7.1 Counterfactual experiments

I investigate how important a consumer's human capital is, in explaining the observed changes in her picture quality, as well as her choices of cameras. We simulate two counterfactual experiments. First, we assume that there is no switching cost in human capital, so that new camera purchases are not associated with human capital losses. Second, we assume that there is no learning by doing, and (because of our normalization) no role of human capital at all. Under each counterfactual scenario, I simulate the choices and outcomes of a group of consumers, who share the same initial condition as that of the data, but evolves according to the counterfactual parameters. I separately plot the counterfactual outcomes against the time that a consumer stays in sample, for the first 5 years, and compare them with simulated choices and outcome under benchmark model estimates. These are shown in Figure 2.12 and 2.13.

I find that when there is no learning, consumers systematically purchase less DSLR cameras and take less pictures. In fact, the sales of advanced cameras is reduced by 20% of its benchmark value. Also, we find that the amount of pictures individuals take is reduced by almost 30%. *Directly*, this is a result of the complementarity of human capital and camera format: that consumers with a larger stock of human capital derive higher payoff from using an advanced camera, and consequently, they are more willing to pay higher prices for that.

In addition, I find that both the demand for advanced cameras and the demand for picture-taking activities are smaller at the very beginning, when the human capital stock is still close to zero. This additional finding speaks for the investment motives in consumers' purchase and picture taking decisions. Specifically, a consumer is willing to purchase a DSLR early, if she foresees that there is product-specific learning by doing, even when her current human capital does not justify it; similarly, she is willing to take more pictures early, beyond her current consumption utility.

Counterfactual outcomes when there is no switching cost is displayed in Figure 2.13. Here, I find that when learning by doing is entirely general, the demand for compact cameras

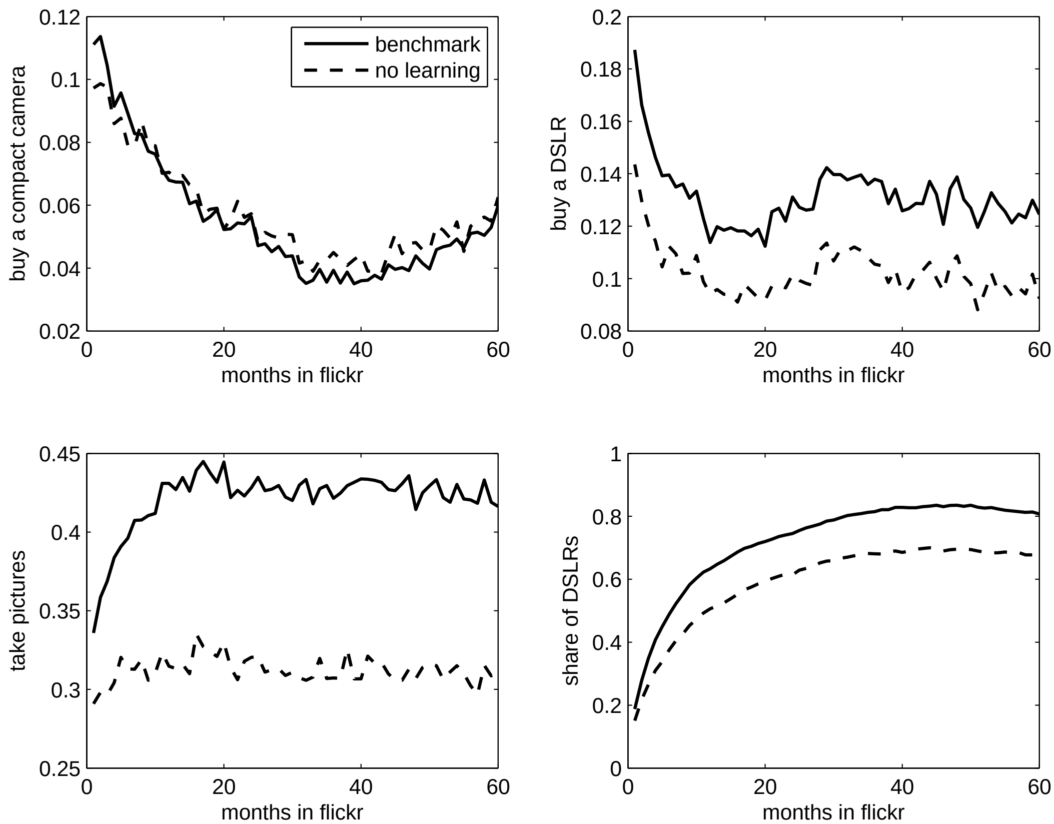


Figure 2.12: Counterfactual outcomes if consumers do not learn

Notes: This figure reports counterfactual outcomes when there is no learning. The four panels report, respectively, choice probabilities of buying a compact camera, buying a DSLR, taking pictures, and finally, share of DSLR holders at a given point in time.

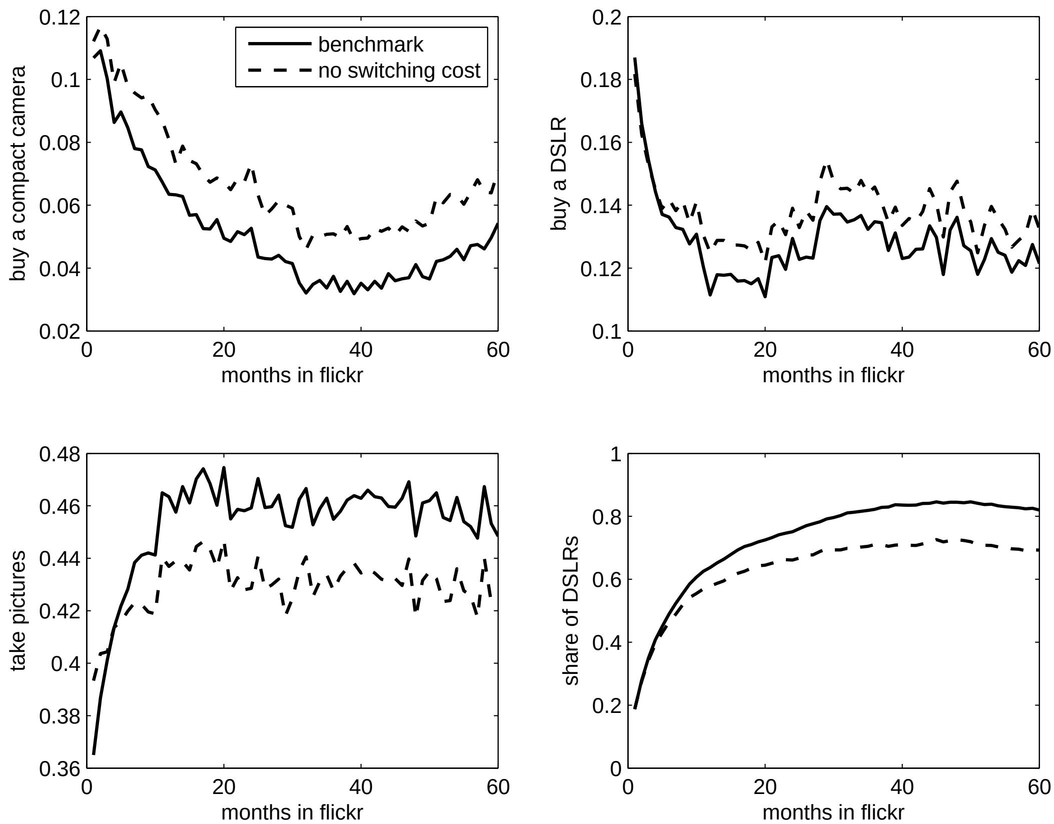


Figure 2.13: Counterfactual outcomes if human capital is fully general

Notes: This figure reports counterfactual outcomes when there is no switching cost. The four panels report, respectively, choice probabilities of buying a compact camera, buying a DSLR, taking pictures, and finally, share of DSLR holders at a given point in time.

increases, by a larger magnitude than the increase in DSLR demand. This is driven by the fact that consumers do not need to purchase an advanced camera early if they expect to learn from it (and benefit from it in the future), if she can "freely" learn from a cheaper alternative. In other words, shutting down switching cost takes away the investment motive for buying a DSLR early. As a result of this, market share of DSLR holders is lower than when there is switching cost; it should be highlighted that although the pattern is similar to that of no learning, the underlying mechanism is very different.

Also, picture-taking frequency decreases by a small amount, when there is no switching cost. The difference goes to the immediately higher incentive to practice when human capital suffers a loss after camera switching.

2.7.2 The demand for human capital

Increase in human capital stock has direct effects on on product-usage utility, and it increases consumer welfare (Michael, 1973). Hence, there is a direct demand for usage experience, which a firm or a third party can provide. In the context of digital cameras, examples of such include free product training,[45] photo contests (to incentivize product usage), and so forth.

I quantify the size of such demand, by giving all consumers one more month of experience, at the very beginning. Specifically, I allow them to all take pictures using their first camera, in period 0. This then increases their beginning human capital stock from period 1, and the added effect decreases over time. I then charge the consumers a fixed monthly fee, and find the amount of the fee such as to bring their expected utility back to the case without training. This utility-equivalent amount of fee is then their willingness to pay, or equivalent variation, of the 1-month added experience.

Formally, denote the expected sum of future utility as

$$V_{it}\left(EV_i, H_{i0}\right) = \mathbb{E}_{it}\left[\max_{\mathbf{A}_{it}} U_i\left(\mathbf{A}_{it}, \mathbf{S}_{it}; EV_i, H_{i0}\right)\right]$$

[45] For example, in India, Vietnam and potentially some other countries, Canon provides free short-lectures for users who just purchased their DSLRs.

as a function of the fixed fee EV_i charged in each period (on top of any price in case of a purchase), and the initial human capital H_{i0}. If the consumer received an education, she draws a signal M_{i0} before period 1. Otherwise, $H_{i0} = 0$ as our normalization. Now, a consumer's willingness to pay for 1-month of usage experience is the amount of fixed fee that equates expected utility without the training and "tuition fee", and the utility with both:

$$V_{it}\left(EV_i, \max\{M_{i0}, 0\}\right) - V_{it}\left(0, 0\right) = 0.$$

And because EV_i is charged in all periods, I present $\frac{EV_i}{1-\delta}$ as the discounted sum of the fixed monthly payment, at $\delta = 0.95$.

Table 2.13 presents the results, separately for consumers from different segments. The first row presents the willingness to pay for a contemporary experience shock, and the second to fourth row present willingness to pay for an experience shock in the past. We find that a one-time experience shock has considerable effect in contemporary consumer value function, because of the steep learning curve when she is relatively inexperienced. In fact, the consumer's willingness to pay for the increase in experience is 1.2 times the magnitude of her expected life-time expenditure in the product category. In addition, the impact on a high-starter is much larger, because she can learn more from the one additional draw. Finally, the one-time experience boost decays slowly over the years, and this means that human capital shock has long-run effects.

2.7.3 Managerial implications

2.7.3.1 Consumer education

We find that consumers have high valuation for even a small increase in human capital. This means that potentially, there is a market for consumer's product-specific education.

A firm or a third-party can organize consumer workshops, and charge full fee for that. In fact, I find that the tuition fee for a one-month New York Institute of Photography online course is \$50,[46] which seems to be an optimization result from a the demand curve, similar

[46] This information is from http://www.nyip.edu/courses/professional-photography, in September 2014.

Table 2.13: Valuation for additional 1 month of experience

	WTP: lo-starters	WTP: hi-starters	WTP: average	lifetime expend.
start of the sample	465.1845	711.3428	554.8836	469.1844
1 year	159.5055	184.6547	168.6698	400.8924
2 years	69.5021	68.9929	69.3165	379.0865
3 years	38.8951	31.6835	36.2672	373.6072
4 years	29.5170	19.8236	25.9848	379.5301

Note: This table reports, for consumers in different segments and with different levels of experience, the amount of compensation that is welfare-equivalent to a counterfactual increase in human capital by 1 month, at the beginning of the sample. I compute the amount of a fixed monthly subsidy, that is equivalent of this policy change – in the sense that it equates the expected sum of future utility for the consumer. I then take discounted sum of this stream of subsidy. For example, the first number reads: for a low-starter consumer at the beginning of the sample, her valuation of the 1- month extra human capital is measured as the utility-equivalent *one-off* tax at 465 dollars. The last column reports the expected discounted lifetime expenditure for consumers with the corresponding experience (without the extra human capital). This number is provided as a benchmark to understand the magnitude of consumer valuation of education.

to one outlined in Table 2.13. There are many examples in the digital camera industry. For example, Canon and some other manufacturers have organized in-store workshops, which are more dedicated to educate brand-specific knowledge, as well as generating traffic in their stores. Also, they have long been sponsors to other third-party held lectures and workshops, which are more dedicated towards general knowledge.

There are two additional caveats from our results. First, consumer education steals the market for entry level products. Since the industry consists of both multi-camera-format firms, as well as firms that dedicate themselves to only compact cameras, these different firms might have different incentives to subsidize consumer knowledge. Specifically, firms that only produce compact cameras might only have incentives to provide firm-specific or format-specific knowledge, while firms that produce both need to carefully weight the revenue gain and losses among their different departments.

Second, we find that a shock in consumer human capital decrease their tendency to purchase compact cameras, but this does not mean that they will immediately buy a DSLR. Sponsoring consumer human capital might prolong their waiting in the learning process (if they expect that more knowledge will come tomorrow), which might cut short-run profit even for a multi-product firm. Therefore, an interesting theme for future study on supply-side provision of consumer knowledge, is the consequence of mis-alligned time horizons.

2.7.3.2 Switching cost and product design

We find that switching cost is increasing in consumer human capital stock, and it is higher when she switches to a very different product. This include switching to a different brand, and switching to a different product format. Therefore, it will become increasingly difficult for a consumer to switch either across brand or across format (especially both).

It should be highlighted that, despite the similarity in magnitude, switching costs across brands and across formats within a brand have very different implications. For two firms of similar market shares, *creating* product design differences might be a way to differentiate their products, and avoid head-to-head competition. For example, Canon lenses zoom in when turned counter-clockwise, while Nikon lenses do so if turned clock-wise. Therefore, for experienced Canon consumers, she would not be as elastic to the prices of Nikon cameras, and therefore both firms can charge higher mark-ups.

Product-format switching cost, on the other hand, might partly be due to product-design limitation, that a firm would wish to eliminate. After all, if a multi-product firm is a monopolist, it might be incentivized to migrate consumers into buying advanced products, and earn higher profits from that (judging by their prices, which are 5-10 times different). However, the story is more complicated if there are multiple firms, and a firm might consider designing its advanced product similarly to the opponent's entry-level product, so as to steal some of the upgrading consumers.

As limitation of this study, we are unable to measure product design features, and hence, unable to quantify how switching cost depends on the "difference" between two cameras (besides brand and format). If we were able to measure this, studying how equilibrium switching cost depends on market structure is an interesting topic, now left for future research.

2.8 Concluding Remarks

This paper quantifies the importance of consumer learning by doing – i.e. accumulation of product-specific human capital through usage – on their demand for advanced products. In the context with entry-level and advanced digital cameras, I measure the returns to consumer experience, via looking at how a homogeneous set of viewers receive a consumer's pictures, taken at different points in time. On the one hand, experience leads to higher utility from product usage (in this case, via higher picture quality). Thus, learning by doing explains a considerable share of the demand for advanced products. On the other hand, I find that up to 26% of a consumer's product-specific human capital is not fully transferable, and this discourages product switching – in particular between brands where product-design differences are greater. As a result, more experienced consumers display greater brand loyalty; and knowing so, even mildly forward-looking consumers consider products across brands to be much more substitutable in the long run, than those within a brand.

The model of consumer learning by doing that is proposed in this paper has great generalizability in home electronics, sports equipment, entertainment, and other categories that require consumer skills to use the products. From a managerial point of view, understanding the evolution of consumer knowledge not only helps understand the evolution of their demand – in particular the migration from entry level to more advanced products, but it also helps understand their tendency to be locked in to products that are similarly designed as their previous ones. Further, because usage experience is desirable on its own, there is demand for supply-side provision of consumer knowledge, such as the firm offering training services, competitions in user content creation, or simply designing products that are easy to use. From the manager's perspective, whether such actions are profitable depends not only on the returns to experience, but also on how widely-applicable the product knowledge is.

The empirical exercise in this paper is done on a relatively small sample, which might be non-representative; however, the difficulty of measuring the returns to experience limits the usage of more standard market share data, used in Song and Chintagunta (2003), Gowrisankaran and Rysman (2012), among others. This is related to the general difficulty of identifying the source of state dependence from choice data alone (Ching et al., 2013), which is beyond the scope of this paper.

2.9 Supporting materials

2.9.1 Returns to experience in photography

Overview This section estimates the returns to experience in photography in reduced form, where the model is flexible enough to allow for deviations from the assumptions used to infer picture quality, in Section 2.3.3. Hence, this section also serves as a robustness check for the picture quality measure.

Specifications I first further assume quadratic specification of picture quality q_{ip} from Equation (2.2):

$$q_{ip} = \theta_1 x_{ip} + \theta_2 x_{ip}^2 + \sum cam_{it} + q_{i0} + \eta_{it} \tag{2.17}$$

which is a quadratic on experience x_{ip}, plus a set of camera fixed effects, individual fixed effect q_{i0}, and a error term η_{it}. I then regress

$$\log\left(views_{ip}\right) = \sum_{t_0,t_1} \Phi_{t_0 t_1} + z_{ip}\psi + \theta_1 x_{ip} + \theta_2 x_{ip}^2 + \sum cam_{it} + q_{i0} + \eta_{it}, \tag{2.18}$$

and the parameters θ_1 and θ_2 capture the returns to experience. Note that this specification shares essentially the assumption I use to infer picture quality, other than the additional quadratic functional form on experience, and the separability in camera dummies.

There are, however, two potential concerns to the assumptions to Equation (2.2). First, the flow of viewers could interact with experience, resulting in heterogeneous display-window effect. In other words, $\Phi_{t_0 t_1}$ might be individual specific. The second concern is associated with the timing of upload, i.e. the user might strategically choose the time to upload a picture based on its quality. Both arguments point to the heterogeneity of the display window dummies.

With this in mind, I also estimate the returns to experience on a more-flexible specification. Although this cannot be used to infer picture quality, it serves as an robustness check. Specifically, I allow for interactions of individual heterogeneity and the display-window ef-

Table 2.14: Returns to experience in photography

	individual fixed effect	individual-display window fixed effect
experience (100 months)	0.641***	0.536***
	(0.012)	(0.015)
experience sq (0000s)	-0.139***	-0.114***
	(0.005)	(0.007)
camera dummies	Yes	Yes
topic dummies	Yes	Yes
upload order	Yes	Yes
display window	Yes	No
months since joined Flickr	Yes	No
number of pics uploaded	Yes	No
Rsq.	0.112	0.008
obs.	1557232	1560224

Note: This table provides reduced form estimates of the returns to experience in photography. The dependent variable is log of cumulative number of views, per picture level. The first column corresponds to Equation (2.18), where we infer the returns to experience using within-individual variation, but adding covariates to control for aggregate calendar time trend in Flickr. The second column reports estimates using within-individual-upload-time variations, based on the specification in Equation (2.19). Measured in *picture quality*, the first specification estimates a 3-year return to experience of 21.3%, or an annualized 6.7%; the second specification estimates a 3-year return of 18.2% – annualized to 5.8%. Within-effect R-squared are provided.

fects, resulting in individual-display-time dummies $\tilde{\Phi}_{i,t_0,t_1}$. Equation (2.18) now becomes:

$$\log\left(views_{ip}\right) = \tilde{z}_{ip}\psi + \theta_1 x_{ip} + \theta_2 x_{ip}^2 + \sum cam_{it} + \tilde{\Phi}_{i,t_0,t_1} + \tilde{\eta}_{it}, \tag{2.19}$$

where $\tilde{\Phi}_{i,t_0,t_1}$ now captures a combined effect of baseline picture quality q_{i0} and individual-specific flow of viewers. We can regress (2.18) controlling for individual-batch fixed effects.

Estimates The first column in Table (2.14) presents the estimation results from Equation (2.18). I find that the returns to experience is positive within sample period, with a decreasing marginal return. This is consistent with the learning curve measures in Shaw and Lazear (2008), Besanko et al. (2010), Levitt et al. (2013), among others. Quantitatively, the *annual* return to experience in the first 3 years amounts to an increase in picture quality, such that it generates 6.7% more views.

I further check the sensitivity of this result to potential endogeneity problems, as discussed. Shown in Column 2, the robust learning speed estimates are not economically

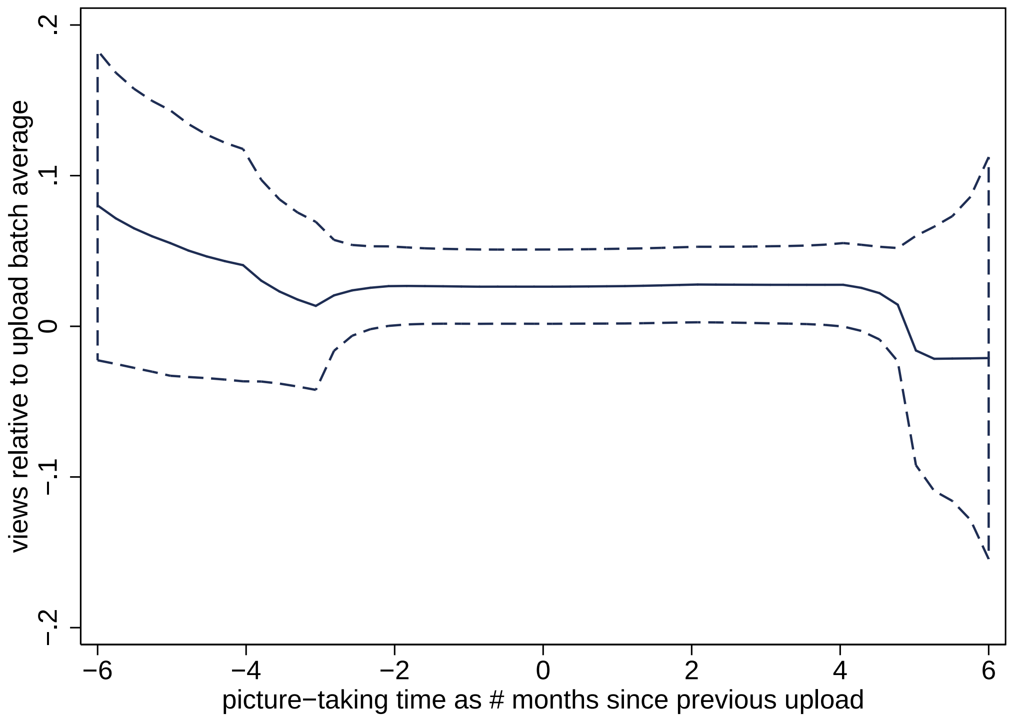


Figure 2.14: Views on pictures uploaded at different points in time

Note: This figure plots views of a picture relative to views specific to a batch, for pictures uploaded at different points in time. Specifically, the horizontal axis is the *picture taking time*, but is normalized as the number of months since the previous upload batch. For example, -2 means that this picture was taken 2 months before an upload, but the picture was only uploaded in the next batch (so this is delayed upload); +2 means that the picture was taken 2 months after the previous upload and is uploaded in the next batch.

different from the benchmark estimates – hence, the inferred learning curve economically robust to the potential concern of endogeneity.

2.9.2 Additional Figures and Tables

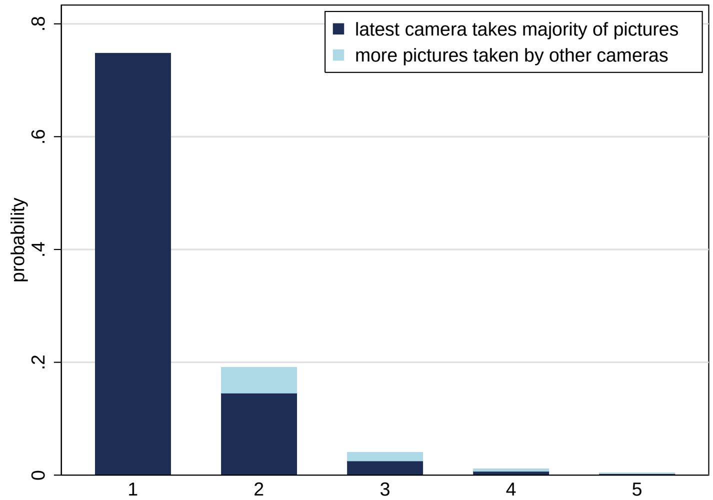


Figure 2.15: *Joint* probability of the number of cameras owned, *and* whether the latest camera takes the most pictures (x-axis: # cameras)

Note: This figure shows the joint probability of the number of cameras owned at a given time, and the incidence that most pictures are taken by the latest camera. A camera is owned at a point in time if I observe at least one picture taken before that, *and* at least one picture taken afterward. By construction, a camera takes all pictures if it is the only "owned" camera – as represented by the dark bar at x axis = 1. When more than one camera is present, I find that in most occasions, the last camera still takes most pictures.

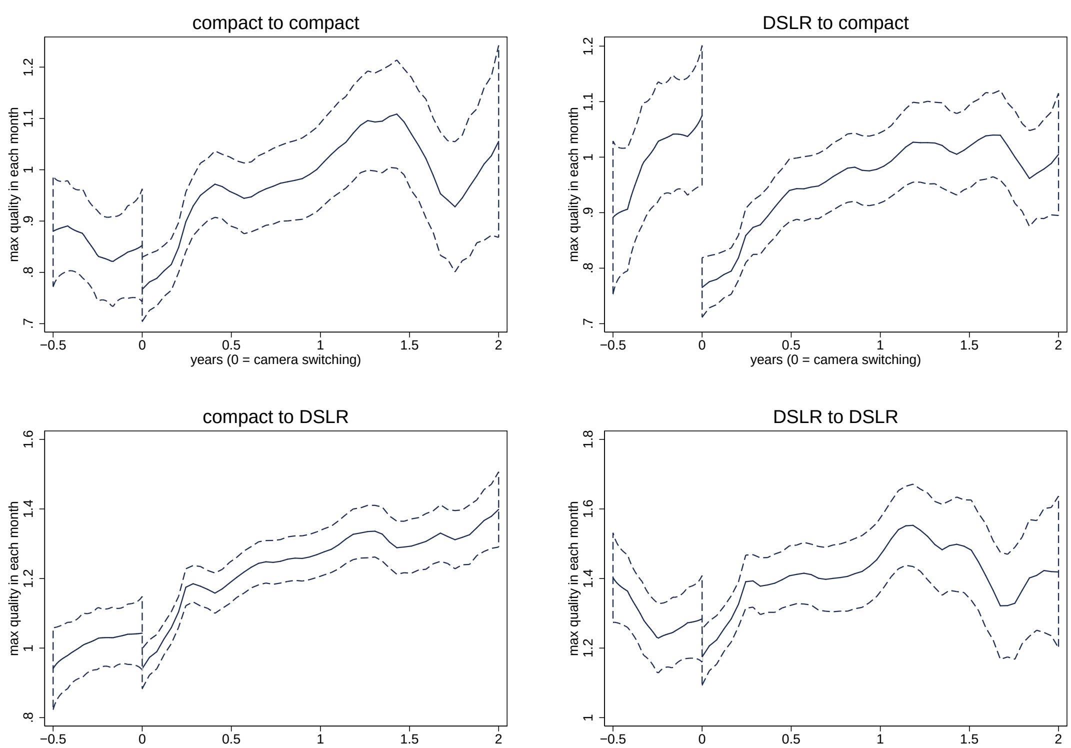


Figure 2.16: Switching cost controlling for camera formats

Notes: These figures present monthly maximum picture quality for each individual, before and after camera switching, conditional on the camera formats before and after. For detailed notes, see Figure 2.3.

Table 2.15: Average short-run elasticities

	(1)	(2)	(3)	(4)	(5)	(6)	no purchase
(1) Canon Compact	-2.6257	0.0317	0.0312	0.0291	0.0304	0.0301	0.0265
(2) Nikon Compact	0.0147	-2.6459	0.0149	0.0139	0.0142	0.0141	0.0116
(3) Other Compact	0.0169	0.0177	-2.6436	0.0159	0.0166	0.0163	0.0136
(4) Canon DSLR	0.0426	0.0433	0.0428	-3.5060	0.0423	0.0428	0.0408
(5) Nikon DSLR	0.0231	0.0240	0.0234	0.0223	-3.5350	0.0232	0.0211
(6) Other DSLR	0.0219	0.0228	0.0222	0.0208	0.0218	-3.5388	0.0182

Note: This table reports short-run elasticities. I compute elasticities by first calculating the implied choice probabilities for each type of consumer, and then the counterfactual choice probabilities when prices for a given brand-format *in a row* are *temporarily* reduced by 10% for the given month. Then, elasticities are computed from the *averaged* choice probabilities. For example, the first row, second column reads: a 10% temporary decrease in the price of Canon compact cameras *decreases* the demand for Nikon compact camera by 0.385%.

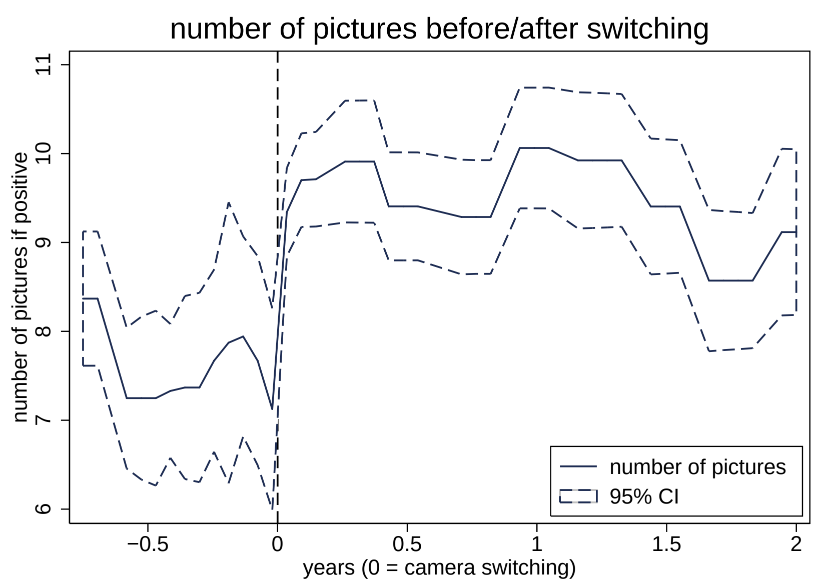


Figure 2.17: Number of pictures before/after camera switching

Notes: This figure presents the number of pictures an individual produces, around the time of camera switching. This is conditional on these pictures eventually being uploaded to Flickr.

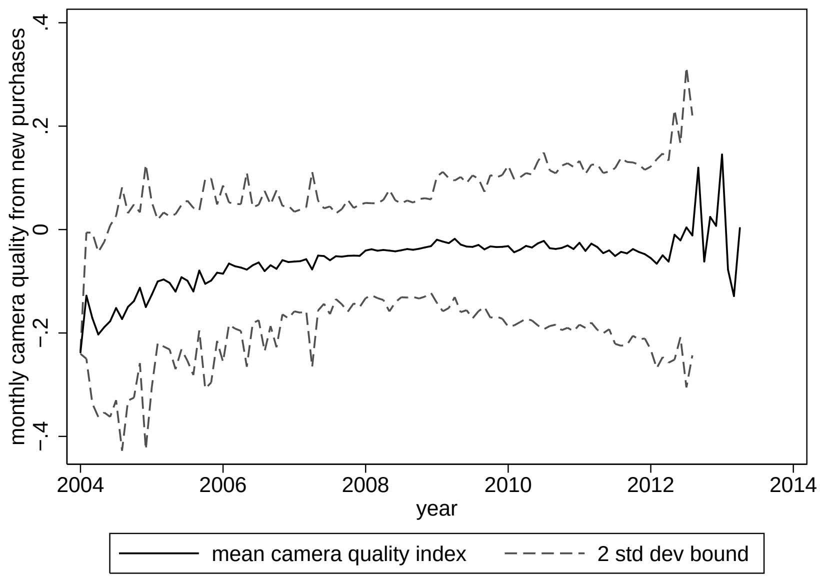


Figure 2.18: Evolution of the projected camera quality distribution

Notes: This figure depicts average camera quality index, $\bar{Q}_j$, among all cameras *adopted* in the same month. That is, the horizontal axis is the year of adoption (in units of 1/12 years). The figure also plots confidence intervals of two standard deviations, which show within-adoption-month variations in $\bar{Q}_j$. Note that these are *not* the standard error (and confidence interval) of the mean estimate, $\mathbb{E}\left[\bar{Q}_j|t\right]$.

Table 2.16: Average long-run elasticities

	(1)	(2)	(3)	(4)	(5)	(6)	no purchase
(1) Canon Compact	-2.8812	-0.0555	-0.1271	0.0336	0.0799	0.0452	0.0269
(2) Nikon Compact	-0.0425	-2.8610	-0.1134	0.0317	-0.0295	-0.0028	0.0124
(3) Other Compact	-0.0716	-0.1034	-2.8752	0.0223	0.0061	-0.0216	0.0148
(4) Canon DSLR	0.8194	-0.3793	-0.0928	-2.7735	-0.7473	-0.3487	0.0399
(5) Nikon DSLR	-0.0881	0.9397	0.1527	-0.3614	-2.7265	0.0032	0.0188
(6) Other DSLR	0.3989	0.4968	0.7359	0.0654	0.2275	-2.7937	0.0142

Note: This table reports long-run elasticities, computed from demand responses to permanent price changes of 10% for a given brand-format of camera. For example, the first row, second column reads: a foreseeable, permanent 10% price decrease for Canon compact cameras *decreases* the demand for Nikon compact cameras, by 0.298%.

Table 2.17: Long-run elasticities without learning by doing

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-2.56	0.12	0.02	0.07	0.05
(2) Nikon Compact	0.08	-2.61	0.04	-0.00	0.03
(4) Canon DSLR	0.27	0.01	-3.49	-0.17	0.04
(5) Nikon DSLR	0.03	0.25	-0.08	-3.43	0.02

Note: This table reports long-run elasticities, assuming no role of picture quality (and its evolution) at all.

Table 2.18: Short-run elasticities without learning by doing

	(1)	(2)	(4)	(5)	no purchase
(1) Canon Compact	-2.58	0.07	0.06	0.07	0.05
(2) Nikon Compact	0.04	-2.61	0.04	0.04	0.03
(4) Canon DSLR	0.05	0.05	-3.50	0.05	0.04
(5) Nikon DSLR	0.03	0.03	0.03	-3.53	0.02

Note: This table reports short-run elasticities, assuming no role of picture quality (and its evolution) at all.

Chapter 3

Quantifying Consumer Consideration Cost

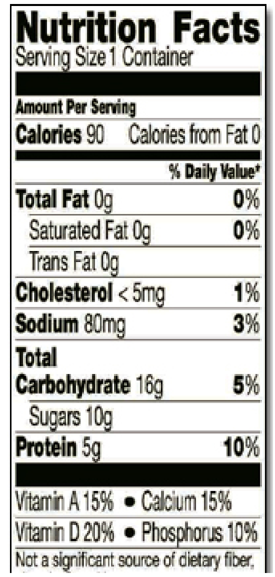
3.1 Motivation

Some product characteristics are more salient than others. For example, even for a regular buyer of yogurt, the product characteristics information provided in a nutrition table, such as the one in Figure 3.1, require quite some mental power to process. In other words, it is much less costly for a consumer to get full knowledge of the brand and price of a product, than to know and understand the nutrition contents – even if she cares about the latter just as much as the former.[1]

We study the implication of a consumer's cost of *processing* the less salient product information, in order to evaluate her own utility from buying a product. We term this concept "costly consideration". It is closely related to "thinking cost" in Shugan (1980), who analytically studies the cost of evaluating products with multiple characteristics. Despite that discussions of consumer decisions under limited processing capabilities go a long way (also see Simon, 1959 and Bettman et al., 1998), the magnitude of such processing cost has not been quantified, especially in field data. This paper quantifies a consumer's consideration cost using a standard consumer scanner data-set.

To fix idea, think about there being only two product characteristics: price and nutrition.

[1] In fact, yogurt manufacturers heavily advertise on nutrition contents, which suggests that consumers are indeed concerned with those.



Nutrition Facts
Serving Size 1 Container

Amount Per Serving	
Calories 90	Calories from Fat 0
	% Daily Value*
Total Fat 0g	**0**%
Saturated Fat 0g	**0**%
Trans Fat 0g	
Cholesterol < 5mg	**1**%
Sodium 80mg	**3**%
Total Carbohydrate 16g	**5**%
Sugars 10g	
Protein 5g	**10**%
Vitamin A 15% • Calcium 15%	
Vitamin D 20% • Phosphorus 10%	
Not a significant source of dietary fiber,	

Figure 3.1: Nutrition table on Yoplait Light yogurt

Notes: This is an example of a nutrition table, on Yoplait Light yogurt from General Mills. Extracted from http://www.yoplait.com/products/yoplait-light in March, 2015.

Price is salient, but nutrition information requires effort to be processed and understood. Because of such processing cost, a consumer would only decide to think about nutrition information when the price information looks "favorable". In other words, she will neglect products that are way beyond her price acceptance range; that is, these products will be out of her consideration set. In addition, she decides to look at, and think about the nutrition table, only if she expects to buy non-trivial amount after doing so. This is because her effort on reading and thinking about the table is a fixed cost, and the expenditure of this cost requires enough purchase quantity (and gain from consuming many units) to be justified.

Formally speaking, the consumer's consideration set formation decision is similar to the conjunctive choice rule – proposed in Shugan (1980) – in the sense that she screens products based on salient information, which is only a subset of relevant product characteristics. In addition, because of the fixed cost nature of consideration cost, products that are screened in must have large enough expected quantity (and total gain from purchase), such that it justifies the fixed cost. Therefore, the price threshold is also a selection rule on (endogenously determined) quantity choice.

To the researcher studying consumer demand, this property is a key identifying restriction, which separates heterogeneity in the consumption preferences, from limited consideration due to such mental cost. Specifically, in presence of consumer preference heterogeneity, some consumers prefer Dannon while other prefer Yoplait, and the two groups might concentrate their expenditure onto different subsets of products. This cross-sectional concentration pattern might have nothing to do with costly consideration. On the other hand, in a panel data structure with price variations, a Dannon lover (who has decreasing marginal utility but no consideration cost) might occasionally find Yoplait price to be marginally favorable, and it pays off to purchase a small amount under the discounted price. However, this is not the case, if the consumer only has a slight preference for Dannon, and previously concentrate on the product *mainly because* she does not want to spend twice the fixed cost. Now, under discounts of Yoplait, she might choose to completely swap Dannon for Yoplait, and purchase a considerable amount the latter. In other words, the joint distribution of product and quantity choice reacts differently to the same price variation, if costly consideration is at play. This identifying strategy, similar to the one in Dehmamy and Otter (2014), is key to quantify consumer consideration cost when the consideration set is unobserved.

In the paper, we first explain a simple model with quantity choice and costly consideration. We analytically and numerically show that the price threshold depends on how much quantity a consumer *would* buy if consideration cost is expended. The model can then derive testable implications, in particular on how much consumer purchase quantity would change at this price threshold; that is, at the maximum price where we observe a purchase. Without imposing any structural models, we test for this using consumer scanner panel data in the yogurt category, from Information Resource Inc. (IRI). We find very large quantity jump when the price just become acceptable for *each* consumer. The magnitude of the quantity jump is not explained by quantity discount (that is, unit price is non-increasing in the quantity one purchases) and indivisible quantity (i.e. there is a minimal quantity that is available), and therefore, strongly speaks for the existence of consumer consideration cost. Our reduced-form evidence can be replicated in some other categories, such as milk, coffee, and frozen pizza.

Next, we structurally quantify the magnitude of consumer consideration cost, by estimating a model of product-quantity choice on micro-level scanner data. To accommodate our identification strategy, our model characterizes consumer choices over multiple products, as well as their quantity choice for each product they purchase. In particular, our model needs to accommodate quantity discontinuities, as this is what we test for in the reduced form. Consequently, standard models for multiple discrete choice (Hendel, 1999; Kim et al., 2002; Dubé, 2004) do not apply in this context, as they rely on the property that optimal quantity choice is everywhere continuous in price. Instead, we explicitly model a two-stage decision problem, in which a consumer first chooses a subset of product to costly consider, and then chooses the quantity within this consideration set. Specifically, we model choices over bundles of products in the first stage, similar to Gentzkow (2007), and then, conditional quantity choices given membership in a consideration set. Because consideration set is unobserved by the researcher, we then integrate quantity choices over all potential consideration sets, similar to Goeree (2008). Although this approach suffers from heavier computational burden (but in our case it is manageable), it allows for choices over multiple product, as well as quantity discontinuity at marginal price, both of which are key to our problem. In addition, we can also flexibly allow for nonlinear prices (quantity discounts) and discrete quantity sets (Allenby et al., 2004). Finally, we abstract away price search as it is non-central to our research question. To accommodate this simplification, we condition consideration and quantity choice on category purchase decisions, both in the model and in the data.[2]

We estimate the model using IRI scanner data in the yogurt category, and find that in this context, consideration cost is between \$3.2 and \$5.2 *per product-trip*. This means that consideration cost is 1.2-1.9 times the magnitude of a consumer's per-trip expenditure on a product. Due to the high fixed cost for doing so, limited consideration is non-negligible for the researcher.

Under costly consideration, an important role of price discount is to incentivize consumers to *think about* a product. As the first implication, we decompose price elasticities into an effect on inclusion of the consideration set, and another effect on quantity choice

[2]Likely, the consumer has traveled to the focal product shelf. We take the argument in Seiler (2013) that these consumers make little effort in searching for price, compared to the consumers who did not buy products in the refrigerated section.

given consideration-set membership. We find that elasticities given consideration-set membership is about 1/3 of the overall price elasticities, which implies that price discount is ineffective on quantity choice once a consumer starts to think about the product. On the contrary, price elasticities on consideration set formation is large, which implies that the primary role of price discount is to incentivize consideration. This means that prices are acting as important drivers of consumer heuristics and attention, and an important motive for planning price discounts should be to penetrate the consideration cost barrier. In addition, higher consideration cost makes thinking about multiple products more *un*attractive, hence intensifies price competition. This implication is opposite to that of price-search cost, *a la* Diamond (1971).

On the other hand, we allow for, and estimate, preference heterogeneity in the consumption utility (taste). As the second implication of this paper, consumers who purchase *despite* high consideration cost are those with high taste. When feature advertising reduces consumer consideration barrier, it attracts consumers who would otherwise *not* purchase – i.e., those with lower taste. Hence, feature advertising "downward-selects" customers with lower taste, who are also more elastic to price discounts. Therefore, setting a product on feature increases the price elasticities for conditional (on purchase) quantity choice, and this explains why price discounts are usually synchronous to feature advertising. In contrast, in conventional models that treat feature and display as persuasive, products under feature have lower elasticity and should be (without further complicating the model) complemented with a price increase.

Our primary contribution to the literature is methodological: we provide a method to quantify consumer's effort in processing the non-salient product characteristics, so as to study the implication of this cost on potential firm strategies. Earlier literature formally characterize such cost (Simon, 1959; Shugan, 1980; Bettman et al., 1998), but do not have to tools to empirically study it. To this end, we use a similar identification strategy to the one in Dehmamy and Otter (2014) – where exclusion restrictions in quantity choice problem help identify consideration cost – and in addition develop testable implications where standard field data could apply. We also develop a model that is consistent with our test, where standard model on multiple discrete-continuous choices (e.g. Kim et al., 2002;

Dubé, 2004) do not apply. At manageably higher computational cost, our model flexibly characterizes the joint quantity choice for each set of products, and is flexible enough to allow for nonlinear prices (quantity discounts) and discrete quantities (Allenby et al., 2004).

In addition, substantively, our empirical results give different insights into the effect of price discount, and informative marketing strategies. On the one hand, we decompose price elasticities due to a change in consideration sets, and those due to changes in consumption decision given consideration. We find the main effect from a price change on consideration set formation, which speaks for the claim that prices alter consumer attention and heuristics, rather than their fully informed decisions. On the other hand, we study feature advertising as an informative tool, and find different interpretation of it compared to the literature, which mostly treat feature as persuasive (or a complementary consumption good *a la* Becker and Murphy, 1993).

3.2 Related literature

This paper empirically studies the role of consumer consideration cost – their cost of processing some (less salient) product information. Simon (1959) and Bettman et al. (1998) discuss potential deviations from standard utility maximization framework, when consumers face such costs. Shugan (1980) formally conceptualize a consumer's thinking cost. In his model, a consumer explicitly makes a trade-off between making mistakes with higher probability, and economizing thinking cost on product characteristics. This leads to a stopping rule of optimal dimensions of characteristics. Our model closely resembles this thought, but we explicitly formalize salient characteristics (which are free to think about), and non-salient ones (which are costly to consider). We also abstract away from dimensions of non-salient characteristics; therefore, our way of modeling thinking cost as a parameter, is an optimality result in the language of Shugan (1980).

Early literature does not empirically study the magnitude as well as implication of consideration cost. We build our identification strategy on Dehmamy and Otter (2014), who emphasize the exclusion restriction of consideration cost (in their context, they call it "attention") in consumer's quantity choice. In their paper, they rely on lab-experimental variations in the number of shelf facing and position of a product, which varies a consumer's choice of attention, but does not directly affect her taste on quantity. Our paper is different from theirs in two ways. First, we develop testable implications where variations in salient product characteristics – such as price – is enough to identify costly consideration. We directly test for the existence of consideration cost using standard consumer panel data, without imposing a structural model. Secondly, corresponding to our identification strategy, we develop a model that explicitly characterizes the identifying variation that we emphasized in reduced form. It turns out that such model is non-standard in the literature, but in our case computationally manageable.

Our paper is also related to the following three branches of literature. First, there is a large literature on limited consideration set due to costly search (Goeree, 2008; Van Nierop et al., 2010; Kim et al., 2010; Seiler, 2013; Pinna and Seiler, 2014). This literature characterizes and quantifies the consumer-side cost for obtaining information – usually, salient information such as price; in contrast, our paper is interested in the consumer's cost of processing (usually non-salient) information. As a result, our paper is interested in a different stage of decision making process, compared to the search literature. Also, non-parametric identification of search cost usually relies on additional information, such as alternative measure on search independently from purchase (Kim et al., 2010), or time resources spent on search (Pinna and Seiler, 2014). Data-sets with such information are usually rare and (close to) cross-sectional. Hence, they do not contain enough longitudinal variation in product characteristics, which are crucial for identification of consideration. Throughout the empirical analysis, we assume that price information is known.[3] In addition, since our key identifying restriction comes from quantity choice response *given that* a consumer reacts to price change, mis-representing when a consumer reacts to price change should not fundamentally change our findings.

Second, since we rely on concentration of expenditure in subsets of products (and variations of concentration due to price changes) to identify consumer consideration cost, we also model multiple product purchase decisions. Hendel (1999); Kim et al. (2002); Dubé

[3] We accommodate this assumption by restricting our sample on consumers who purchase at least one yogurt of any kind in the given trip. Implicitly, this assumes that she has "scanned through" all the price tags.

(2004) model a consumer's product and quantity choice, and make simplifying assumptions to isolate the choice problems of different products. While this approach eases computation burden, from 2^J to J potential options (J is the number of products), the isolation of quantity decisions across products simplifies away the intense competition among them, for consideration set membership. This is a crucial managerial problem that cannot be ignored. In our paper, competition for consideration is modeled by a separate consideration stage; and is supported by the observed discontinuous quantity jump at certain price threshold, in response to a marginal change in price.

Finally, our paper is related to earlier empirical work on role of marketing mix (such as feature advertising) using consumer package goods data. To our knowledge, many of earlier works (for example, Chintagunta, 1993) use a simple reduced-form model to capture the role of feature and display – as complementary consumption goods that directly affect utility. In our paper, we model feature and display as informative, and emphasize that treating them as pure consumption goods might mis-lead one's understanding of their effects. For example, treating feature advertising as complementary goods would conclude that featuring a product reduces price elasticities; this would then lead to the conclusion that feature and price discounts should be asynchronous. However, our model predicts that feature advertising invites consideration of those with lower brand preferences (i.e. those who would otherwise not purchase). Therefore, featuring leads to overall higher elasticity and justifies a simultaneous price discount.

The rest of this paper is organized as follows. Section 3.3 presents the data. Section 3.4 then discusses the model. First, we outline an illustrative model, which is simple enough to provide analytical as well as numerical solutions. This then generates key testable implications for the existence of costly consideration. We then test for this without any structural model. Then, Section 3.5 parametrizes the model and discusses estimation details. Section 3.6, 3.7 and 3.8 discusses, respectively, parameter estimates, implied prices elasticities and its decomposition, and implications about feature advertising. Finally, Section 3.9 concludes.

3.3 Data

3.3.1 Construction

We use the Behavioral Scan panel data from Information Resource Inc. (IRI) Academic Data Set (Bronnenberg et al., 2008), in the years 2001 to 2007. We focus on the yogurt and yogurt drink categories. A "store visit" is recorded when a household purchases yogurt or yogurt drink in a specific trip – and we assume that the household has traveled to the product shelf of interest. The data records, at the SKU level, the number of units the individual purchased in a given store-week, the total amount paid for the purchase, store level weekly data on the total units sold and revenue received on the given SKU, as well as product characteristics – importantly package size.

At the SKU level, price is defined as the outlet level revenue divided by the outlet level units sold. For price changes (discounts), we define it as the percentage change relative to the maximum price in the past 4 weeks – the underlying rationale being that there are many weekly discounts where prices drop in one week but resume in the neighboring ones.[4] The data-set also records whether the product is on feature advertising or in-store display, or both.

Next, we aggregate the SKU level data. We define a "product" as one with a specific name recorded by the data, regardless of the flavor or package size. For example, "General Mills Columbo Light" is considered as a product, where "General Mills Columbo Light in berry flavor in 8 oz" is a distinct SKU.[5]

We consider the same product with different package sizes as different quantity options of a homogeneous product. To this end, we find the minimum *available* package size of a product, and define "equivalent units" as total purchased volume divided by the minimum package size. For example, for a product with the minimum package size of 8 oz, an individual who purchased 1 unit of 8 oz, and 4 units of 10 oz, is considered to have purchased *6*

[4]One needs to assume that prices do not temporarily increase beyond its regular level. We find some cases with price increase, yet such events occur much less frequently compared to price decrease.

[5]As a robustness check, we alternatively defined a "product" as a product name - flavor category combination, and the essential qualitative reduced-form evidence remain robust.

Table 3.1: Demographics

	Mean	Median	StDev
family size	2.5	2.0	1.3
age of household head	54.5	60.0	14.8
household annual income	46687.0	40000.0	28119.7
obs.	5738	5738	5738

Notes: This table reports household-level summary statistics in the year 2005. Annual income is nominal.

equivalent units. Since few consumers bought non-integer equivalent units,[6] we can characterize quantity choice as discrete (in integer units), and price as a step function of quantity. This captures large-quantity discounts which is frequently observed in this product category.

Finally, we average prices, feature advertising and in-store display into product-quantity level. To do so, we take purchase quantity-weighted average of each variable, among all SKUs for a given product, in a given store and week. For binary variables such as feature and display, we record the associated probability on the product level.

3.3.2 Summary statistics

3.3.2.1 Demographics

There are 8,397 households in the sampling period. Taking a cross-section in the year 2005 (which consists of 5,738 unique households), we find that these households have an average size of 2.5 members, an average age of 54.5 years, and an annual income of $46,687.

3.3.2.2 Trips

Table 3.2 summarizes the duration for each consumer-retailer combination to be in the sample periods, and the number of weeks within that duration when we observe purchase. Overall, a household is present in a duration of 56.4 weeks between the first and the last trip to the same retailer (with purchase of yogurts), within which 10.4 weeks are associated with

[6] In the full sample, 11% of all positive quantity choices involve non-integer equivalent units; yet, only 6% are associated with choices between 0.2-0.8 units. Therefore, even though rounding might introduce non-negligible errors, only a small portion of the sample (0.2% of the overall sample, 6% of the sample with positive quantity) is affected.

Table 3.2: Trips

	Mean	Median	StDev
duration (weeks)	56.4	47.0	53.3
purchase incidence	10.4	4.0	18.2
incidence (top 10 products)	5.7	2.0	11.6
incidence (top 3 products)	2.5	1.0	6.5
obs.	19761	19761	19761

Notes: This table reports summary statistics on the number of trips, per household-retailer, conditional on purchase (of a certain range of products).

yogurt purchases. Within those 10.4 weeks, 5.7 weeks are associated with purchase of the top 10 products, and 2.5 weeks are with the top 3 products.

We drop weeks without yogurt purchase, for a given household and store, and only focus on the product choice conditional on buying at least one product within the category. The purpose of doing so is to condition on price awareness – essentially, we assume that the household has at least "glanced through" all the price tags on the shelf, and hence the price information is free. Thereby, we avoid having to model a costly search process on top of our model.

3.3.2.3 Products, prices and concentration

In the sampling periods, there are 84 distinct products. We compute in-sample market share, based on the shares of total consumer expenditure in yogurt, and find that concentration is only moderate: overall, the top 3 products take up 30% of the market, while the top 20 occupy 82%.

Among the top 20 products, average per-volume price is 0.14 dollar/oz, with a standard deviation of 0.06. For all products the mean (standard deviation) of unit price becomes 0.12 (0.04).

3.3.2.4 Variety and quantity

In sharp contrast to market concentration, we find that within-trip expenditure is heavily concentrated. Among all shopping trips (household-week-store combination), consumers do purchase considerable quantities: 74% are involved with more than one (equivalent) unit

Table 3.3: Variety and quantity

	nr. prod. 1	2	3	4	5	6	Total
1 unit	24.46	0.00	0.00	0.00	0.00	0.00	24.46
2 units	14.72	2.72	0.00	0.00	0.00	0.00	17.44
3 units	8.27	2.12	0.29	0.00	0.00	0.00	10.68
4 units	9.74	2.21	0.32	0.03	0.00	0.00	12.31
5 units	6.12	1.97	0.36	0.04	0.00	0.00	8.50
6 units	6.81	1.96	0.36	0.05	0.01	0.00	9.18
7 units	1.48	1.23	0.31	0.07	0.01	0.00	3.10
8 units	1.83	0.93	0.30	0.05	0.01	0.00	3.12
9 units	0.72	0.59	0.22	0.05	0.01	0.00	1.59
10+ units	5.59	2.73	0.94	0.29	0.06	0.02	9.64
Total	79.73	16.46	3.09	0.60	0.10	0.02	100.00

Notes: Reports percentage share of observations for a given variety-quantity combination.

of purchase, and 23% more than 5 units. However, consumers are not willing to spread the expenditure across different varieties: 97% purchases *one or two* products.

3.3.2.5 Discounts, feature and display

Feature and display are rare in the yogurt category. On the product-store-week level, we find that there are only 7.1% of all observations with *more than half* of the SKUs for a given product are on feature.[7] Displays are even fewer – 1.5% are on display.

Conditional on whether a product is on feature or display, we can then plot the distribution of price changes, shown in Figure 3.2. Discounts are frequently aligned with feature or display: 80% of the products on feature or display are on a discount with no less than 5% price drop. On the contrary, when there is no feature or display, only 8% of the products are on a 5%-or-more discount.

Further, conditional on being on feature or display, the discount distribution has a mode of 40%. This suggests an unwillingness to set shallow discounts. Section 3.4.5 provides a justification to this.

[7] To be precise, the percentage of feature is calculated as a sales volume weighted average.

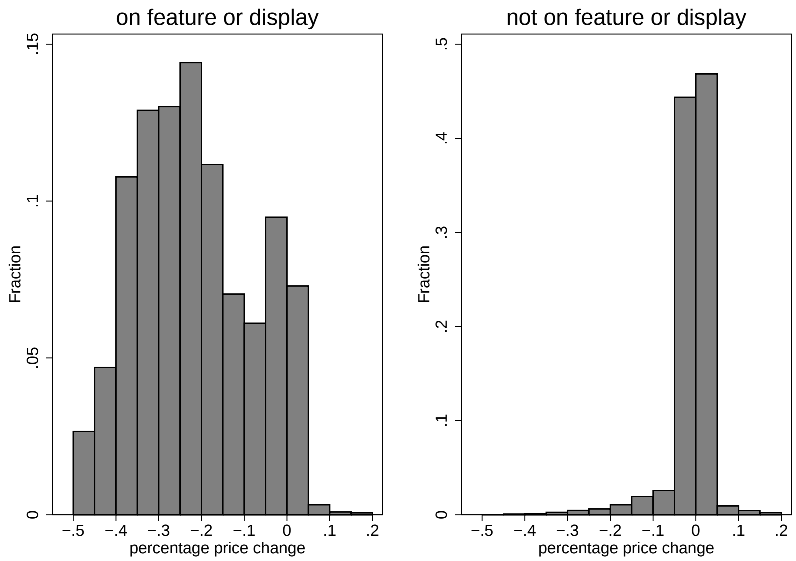


Figure 3.2: Distribution of price changes

Notes: This figure plots the distribution of percentage price changes. Base price is defined as the maximum price in the past 4 weeks. Feature and display are defined as at least *half* the SKUs for a given product are on feature or display, respectively.

3.4 Model setup and testable implications

3.4.1 Overview

In this section, we outline the general model without specifying the functional form. Then, to illustrate the key idea, we provide a numerical example using a simplified version of this model, which clearly provides testable implications. We then provide supporting evidence, directly from the data.

3.4.2 General setup

An individual i, determined to purchase at least one unit of yogurt, travels to the refrigerated product shelf in trip t, where J different yogurt varieties, and an outside option ("other varieties"), are present. The individual knows the price of each product, but needs to incur some consideration cost, to include *each* of the J products into her consideration set $K \subset J$.[8] She then chooses quantity q_{ijt} for each product j in the consideration set.

Denote the $(J \times 1)$ vector of purchase quantity $\mathbf{q}_{it}$. The consumer's full utility can be specified in the following form:

$$u_{it}\left(\mathbf{q}_{it}, K_{it}\right) = \underbrace{c_{it}\left(\mathbf{q}_{it}\right)}_{\text{consumption utility}} - \underbrace{F_{it}\left(K_{it}\right)}_{\text{fixed cost}} + \underbrace{\lambda_i m_{it}}_{\text{money}} \tag{3.1}$$

where $c_{it}(.)$ is a random consumption utility function, which takes as arguments purchase quantity $\mathbf{q}_{it}$, as well as consumption utility shocks from trip t, $F_{it}(.)$ is a random fixed cost function, which is dependent on consideration set K_{it} as well as t-specific shocks. Finally, m_{it} is the money left from the trip, that can be spent on the outside options, which brings utility $\lambda_i m_{it}$.

The consumer maximizes her utility (3.1) subject to the budget constraint

$$\sum_{j \in K} p_{jt}\left(q_{ijt}\right) \cdot q_{ijt} \leq m_{it}, \tag{3.2}$$

where we do not impose constant unit price, but rather specify the unit prices $p_{jt}\left(q_{ijt}\right)$ as functions of quantity. The budget constraint then incorporates prices into the consumer maximization problem.

3.4.3 An illustrative model

When the consumption utility is continuous and the fixed cost is strictly positive, purchasing a very small (i.e. close to zero) amount will not justify the effort spent in evaluating the product. For the consumer, she needs to purchase a minimum quantity to justify the consideration cost. This then implies that, when price of a product drops to the point when a consumer finds it optimal to *start purchasing* it – we call this price the "threshold price" – she always purchases above a minimum quantity threshold. Therefore, when the fixed cost

[8]Depending on the context, J denotes either the number of products or the full set of products.

is positive, the observed quantity choice of a given consumer is strictly positive (away from zero) at the threshold price, but (by construction) zero when the price is above the threshold price. This is to say, quantity jump at the threshold price uniquely identifies consumer fixed cost.

To illustrate this, we greatly simplify the model in Section 3.4.2. Despite that this simplification brings the model far from realistic, it gives an explicit solution to the individual demand function, which serves as a good concrete example. Note that the simplification in this section does not correspond to our empirical exercise.

Suppose that there is only one product of interest. In the utility specification (3.1), let the consumption utility be a quadratic function: $c_{it}(q_{it}) = q_{it} - \frac{\beta_i}{2} q_{it}^2$, and a random fixed cost incurs whenever the consumer considers the product: $F_{it}(K_{it}) = f_{it} \cdot (||K|| = 1)$. Also, we assume that the unit price is constant in quantity; and we normalize the marginal utility to money, λ_i, to 1. Finally, note that we have assumed away randomness in the consumption utility $v_{it}(.)$, so that the consumer knows the exact quantity she will purchase. The consumer problem then reduces to

$$\max_{q_{it}} q_{it} - \frac{\beta_i}{2} q_{it}^2 - p_t \cdot q_{it} - f_{it} \cdot \mathbf{1}(q_i > 0).$$

Note that, in this case, a non-empty consideration set is equivalent to purchasing positive quantity.

The consumer solves the problem backwards: she first determines the optimal quantity subject to consideration, and then chooses whether it is optimal to consider the product. Her optimal quantity given consideration is determined by the first order condition:

$$1 - \beta_i q_{it} - p_t = 0$$

which gives $q_{it}^{foc} = \beta_i^{-1}(1 - p_t)$.

Her consideration decision then involves comparing utility from consuming q_i^{foc} and zero. This is to say, $q_i^* > 0$ if $u_i\left(q_i^{foc}\right) > 0$, which yields

$$q_{it}^* = \begin{cases} \beta_i^{-1}(1-p_t) & \text{if } \beta_i < \frac{1-p_t^2}{2f_{it}} \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

3.4.4 Detecting demand jumps at the threshold price

Note that in the above example, the "threshold price" is determined by the price at $\beta_i = \frac{1-\bar{p}_{it}^2}{2f_{it}}$, or $\bar{p}_{it} = \sqrt{1-2\beta_i f_{it}}$. Since this is the highest price that this consumer will accept, the lowest quantity she will ever purchase (other than zero) is

$$\bar{q}_{it} = q_{it}^*(\bar{p}_{it}) = \frac{1-\sqrt{1-2\beta_i f_{it}}}{\beta_i}.$$

That is, given $f_{it} \gg 0$, when price crosses the threshold, the observed purchase quantity will jump between zero and $\bar{q}_{it} \gg 0$. In other words, the researcher can observe a quantity jump even when consumption utility itself is continuous. Figure 3.3 graphically illustrates this.

If the researcher knows the threshold price $\bar{p}_{it}$ for each consumer-trip, she can then test whether the consumer fixed cost f_{it} is positive, by testing whether the quantity at a price *slightly* below $\bar{p}_{it}$ – i.e. the threshold quantity $\bar{q}_{it}$ – is significantly different from zero.

We construct the threshold price $\bar{p}_{it}$ as the maximum price that we observe individual i making a purchase, *around* period t. Specifically, we focus on 4 weeks before the focal period t, and divide price reductions relative to the 4-week maximum price into 5-cent grids. However, because yogurt units are indivisible, given that the consumer purchases at $\bar{p}_{it}$, her quantity choice should by construction be no smaller than 1. Therefore, we measure quantity in multiples of the minimum *available* package size, so that quantity 1 is always feasible for the consumer. Then, we instead test whether quantity choice at the threshold price is greater than 1.

This results in Figure 3.4, which shows that the average purchase quantity at the threshold price, given that the consumer purchases, is between 3 and 4 times the minimal *available*

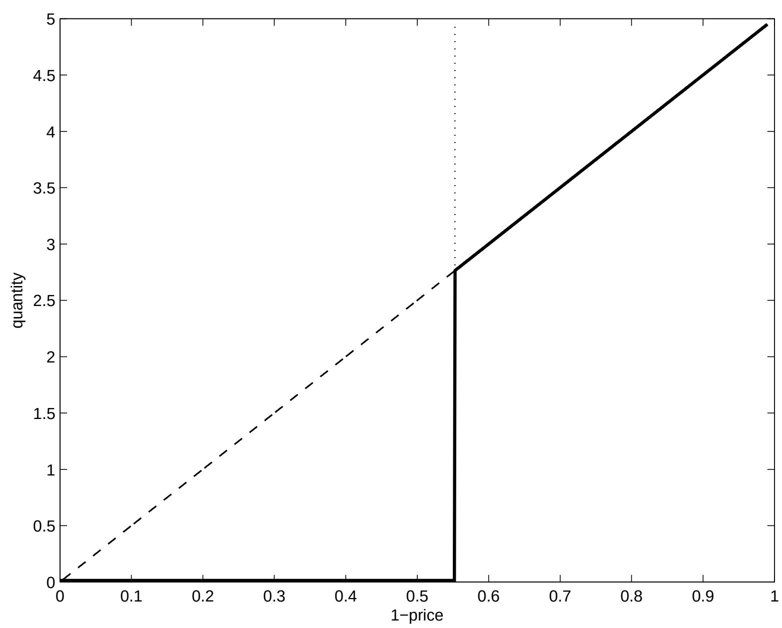


Figure 3.3: Quantity choice as a function of price

Notes: The solid curve illustrates the demand schedule implied by the simple toy model. The dotted line is the "threshold price" $\sqrt{1-2\beta_i f_{it}}$ (she is indifferent between purchasing or not), and the dashed line is the optimal quantity given purchase, q_{it}^{foc}, in the cases when she did not choose to purchase. Specifically, $F = 2$ and $\beta_i = 0.2$.

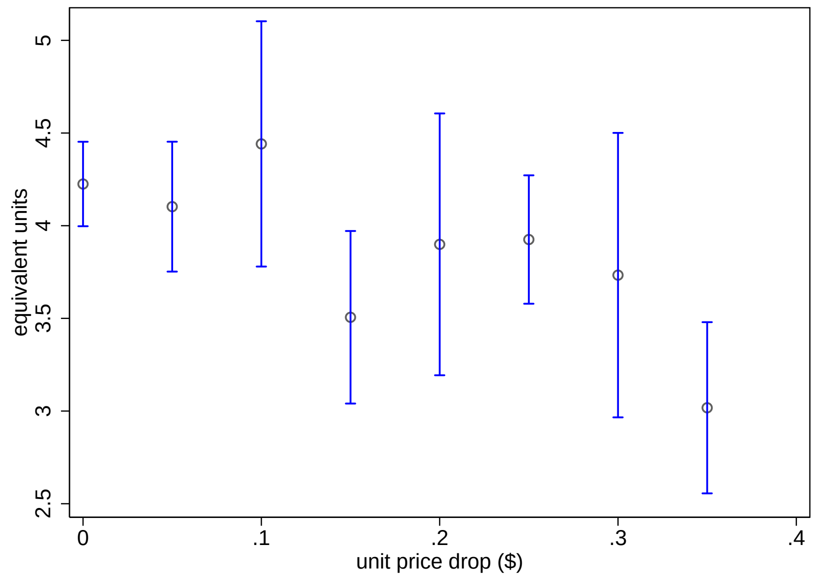


Figure 3.4: Quantity response to price for the marginal consumer

Notes: X-axis is the price reduction (the difference between regular price and current price), at which the consumer would *not* purchase the product, but would purchase at a slightly larger discount. Y-axis is the total number of equivalent units the consumer purchases for the given product.

package size. This shows that a consumer will purchase a non-negligible amount even at marginal changes in price, which implies some discontinuity in the demand function. Under our model specification, this jump in quantity identifies the magnitude of consumer fixed cost.

In addition, for the consumers who do not respond to sizable price reductions, their purchase quantity is slightly lower. This reflects heterogeneity in consumer tastes, which causes the marginal consumers at a lower price threshold to purchase lower quantity.

3.4.5 Aggregate price response across consumers

When the researcher does not observe $\bar{p}_{it}$, quantity response discontinuity is smoothed away because of heterogeneity in the threshold price. However, since the threshold price follows a distribution generated by preference heterogeneity, the market average demand function is the average of individual demand, such as the ones defined in (3.3), but with heterogeneity in β_i. We numerically generate one possible average demand schedule in Figure 3.5. Note

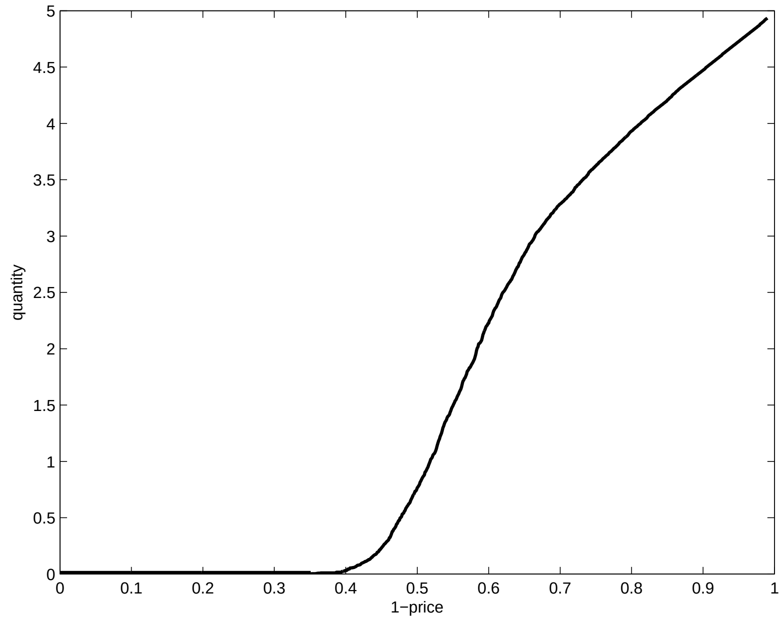


Figure 3.5: Downward-selection of marginal consumers

Notes: This figure is generated by averaging across different curves in Figure 3.3, but with different β_i. Specifically, we assume $\log(\beta_i)$ to be Normal with mean $\log(0.2)$ and variance 0.1. Other parameters follow the previous figure.

that the average quantity response to price change displays an S-shape, in which the steepest part reflects aggregation across different thresholds.

We replicate this shape using our data. Specifically, conditional on that the consumer faces a discount, we look at the *deviation* from her average purchase quantity among her trips in the same store, among the past four weeks when there is no discount.[9] As shown in Figure 3.6, the result is an S-shaped average quantity response function, which indicates lack of response to small price changes. This provides further support to the model without having to assume the location of price thresholds. Related to our evidence, Van Heerde et al. (2001) finds similar evidence on the aggregate sales data.

[9]We focus on deviation to take out individual and product fixed effects.

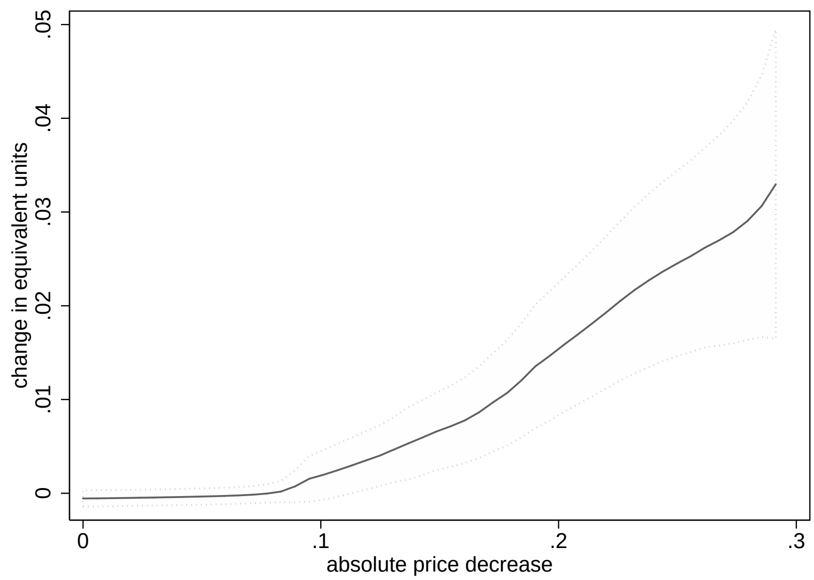


Figure 3.6: Response to discount

Notes: The Y-axis plots deviations from the average equivalent units purchased in the same quarter, under no price discounts; while the X-axis plots the *current* price discount percentage. This figure level of individual-store-week-product. Vertical axis is defined as deviation in quantity (units) from the average equivalent units purchased when there is no discount. Discount (percentage price decrease) is defined relative to the *maximum* price in the past 4 weeks.

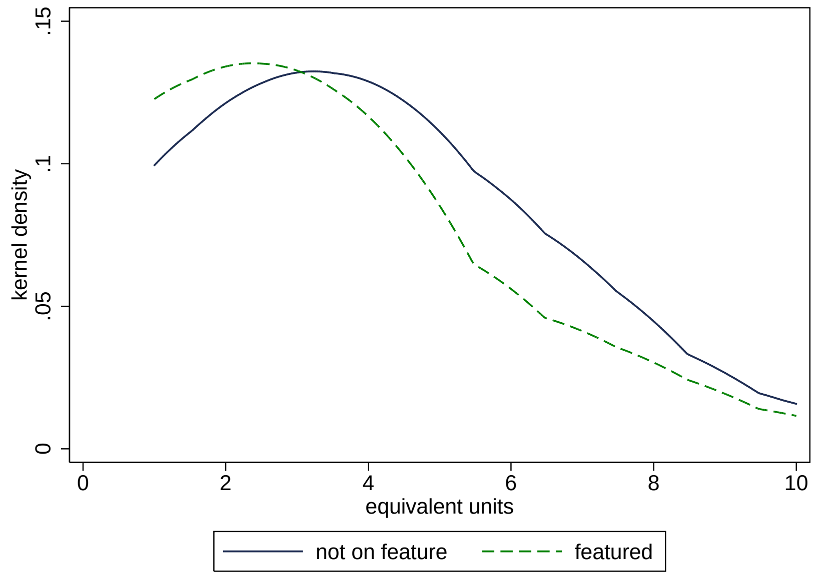


Figure 3.7: Conditional quantity distributions: featured or not

Notes: Kernel density plots of observed quantity distributions for products sold not on feature, and for products sold on feature (dashed). This is the distribution across all consumers, stores, dates and products. Bandwidth is set to 2 for smoothness. the only difference is that in this figure, we focus on the consumers who have never purchased without feature before.

3.4.6 Downward-selection of marginal consumers

Finally, with heterogeneity in β_i, consumers who *only* purchase under lower fixed cost – presumably during feature advertising – are consumers with lower consumption utility. Therefore, a further implication by the model is that purchase quantities for consumers who only purchase under feature is lower than those of the regular customers.

We test for this implication by plotting the observed density function of purchase quantity, given that it is positive, and conditional on whether the product is on feature or not. Shown in Figure 3.7, for products sold on feature, the mode quantity choice for consumers *who purchase* is 3 equivalent units, compared to the 4 units for products not on feature. This suggests that, despite selling more in total, products on feature sell less *per consumer*, reflecting selection of the group with lower consumption utility than a regular customer, who would purchase less.

3.5 Full structural model and implementation

3.5.1 Overview

Recall that in the general model in Section 3.4.2, the consumer i maximizes her utility subject to the budget constrain in 3.2:

$$u_{it}(\mathbf{q}_{it}, K_{it}) = \underbrace{c_{it}(\mathbf{q}_{it})}_{\text{consumption utility}} - \underbrace{F_{it}(K_{it})}_{\text{fixed cost}} + \underbrace{\lambda_i m_{it}}_{\text{money}},$$

which is defined on a vector of quantity $\mathbf{q}_{it}$, and fixed cost as a function of consideration set $K_{it} \subset J$. In our empirical implementation, we limit dimensionality by restricting the total number of products in the consideration set to be at most 2, i.e. $||K|| \leq 2$. In fact, as shown in Table 3.4, there are only 18 out of more than 16,000 observations, where an individual violates this restriction and purchased more than 3. Hence, imposing the consideration set size at 2 is not far from being realistic.

With this restriction, we can specify consumption utility and consideration cost, both as functions of membership and quantities of two products. This section provides details on parametrization of each part of the model, solution of the optimal choice rules, and implementation in estimation. Finally, this section also documents our choice of a sub-sample in estimation, to alleviate computation burden that would be prohibitive in the full sample.

3.5.2 Parametrization

3.5.2.1 Consumption utility

Denote the *purchase set* as $L_{it} = \{j | q_{ijt} > 0\}$. L_{it} is a subset of K_{it}, which is in turn a subset of all products J. Because we restrict the size of consideration set, we can denote

$L_{it} = \{k,l\}$.[10]

We specify the (random) consumption utility as

$$c_{it}(\mathbf{q}_{it}) = \sum_{j\in L_{it}} \tilde{\beta}_{ij} \log(q_{ijt}+1) + \gamma_i \prod_{j\in L_{it}} \log(q_{ijt}+1) \cdot \mathbf{1}(||L_{it}||=2) + \mu_{it}(q_{ikt}, q_{ilt})$$

where the consumption sub-utility is specified in log, and defined on *discrete* quantities $q_{ijt} \in \{0,1,...,\bar{q}\}$.[11] Coefficients $\tilde{\beta}_{ij}$ and γ_i capture the shape of marginal utility in consumption: the $\tilde{\beta}_{ij}$'s capture marginal utility to *percentage* increase in quantity for each product j, while γ_i capture the interaction between percentage increases in quantities between two products in the choice set. When the choice set is singleton, the interaction effect is set to zero.

Also, the consumer derives random consumption utility shocks, $\mu_{it}(q_{ikt}, q_{ilt})$, unobserved by the researcher. One could consider $\mu_{it}(q_{ikt}, q_{ilt})$ as information (or utility shock) that favors, or opposes, purchasing a specific quantity combination. We assume, for a particular combination of quantity (q_1, q_2), $\mu_{it}(q_1, q_2)/\kappa$ is type-1 extreme value distributed, , where κ is a scale coefficient.

Finally, we allow correlation in individual preferences across brands, by imposing $\tilde{\beta}_{ij}$ as a function of observed product characteristics:

$$\tilde{\beta}_{ij} = x_j^{'} \beta_i$$

where β_i is a vector of random coefficients on characteristics, such as brand dummies and low sugar ("light"). Although each dimension of β_i is independent of each other, this introduces dependence within an individual, in $\tilde{\beta}_{ij}$ across product j's.

[10]For notational simplicity, we use the same notation to denote singleton sets – in which case one can think of product l is "0", and the singleton set can be denoted as $L_{it} = \{k\} = \{k,0\}$.

[11]The consumption utility function takes the form similar to Kim et al. (2002) and Dehmamy and Otter (2014), but imposes curvature of the consumption utility. Yet, it allows for consumption utility to interact between products. One other benefit of using the log specification, is that the interpretation each unit increase in the sub-utility is clear.

3.5.2.2 Budget constraint

m_{it} characterizes the attractiveness of the outside option, or "money". Prices affect decisions via a budget constraint, characterized in Equation 3.2. Note that we allow the per-product expenditure $p_{jt}(q_{ijt}) \cdot q_{ijt}$ to be non-linear in quantity, to capture the potential quantity discounts that consumers could benefit from, by buying large quantities. Operationally, $p_{jt}(q_{ijt})$ is the *lowest*, per-unit price one could get when choosing quantity q_{ijt}. λ_i is the price coefficient when substituting the budget constraint to the direct utility function.

3.5.2.3 Fixed cost

The consumer also "pays for" the consideration cost, $F_{it}(K_{it})$, as a function of her consideration set. We parametrize it as

$$F_i(K_{it}, \mathbf{A}_{it}) = \sum_{j \in K_{it}} (f_{ij} + \Delta f_A \cdot \mathbf{1}(A_{ijt} = 1)) + \Delta f_2 \cdot \mathbf{1}(||K_{it}|| = 2) - \varepsilon_{iKt}$$

where $\mathbf{A}_{it}$ is a vector that indicates whether each product is on feature advertising, f_{ij} denotes the baseline consideration cost for each product j for individual i, Δf_A is the increase (negative means reduction) in fixed cost when a product is on feature, and Δf_2 is the additional total consideration cost when considering two products.

The consumer also incurs an unobserved (by the researcher), set-specific utility shock ε_{iKt}, for considering product set $K = K_{it}$. ε_{iKt} are independent type-1 extreme value random variables, across individual, trip and all potential sets $K \subset J$. In addition, ε_{iKt}'s are independent of $\mu_{it}(q_{ikt}, q_{ilt})$.

3.5.3 Solution of optimal choice rules

3.5.3.1 Second stage decisions

Both the researcher and the consumer solve the decision process backward. In the second stage, conditional on the consideration set K_{it}, the individual maximizes utility given consumption utility shock $\mu_{it}(.)$, and chooses the quantity combination. In other words, she chooses (q_{ikt}, q_{ilt}) given $K_{it} = \{k,l\}$ ($l = 0$ in case of a single-product consideration set), to maximize utility (3.1) subject to the budget constraint (3.2).

Substitute (3.2) into (3.1), and we have the indirect utility at the second stage:

$$\begin{aligned} w_{it}\left(q_{ijt}, q_{ikt}; \mathbf{A}_{it}\right) &= \sum_{j=k,l} \left(\mathbf{x}_j^{'}\beta_i \cdot \log\left(q_{ijt}+1\right) - \lambda_i \cdot p_{jt}\left(q_{ijt}\right) \cdot q_{ijt}\right) \\ &\quad +\gamma_i \prod_{j=k,l} \log\left(q_{ijt}+1\right) \cdot \mathbf{1}\left(||L_{it}||=2\right) + \mu_{it}\left(q_{ikt}, q_{ilt}\right) - F\left(\{k,l\}, \mathbf{A}_{it}\right) + \varepsilon_{iKt}. \end{aligned}$$

Denote $\bar{w}_{it}\left(q_{ikt}, q_{ilt}\right) = \sum_{j=k,l}\left(\mathbf{x}_j^{'}\beta_i \cdot \log\left(q_{ijt}+1\right) - \lambda_i \cdot \log\left(q_{ijt}+1\right) \cdot q_{ijt}\right) + \gamma_i \prod_{j=k,l} \log\left(q_{ijt}+1\right) \cdot \mathbf{1}\left(||L_{it}||=2\right)$, and because the consideration cost with shock ε are irrelevant for the second stage decisions, we have the standard Logit choice probability

$$\Pr\left(q_{ijt}, q_{ikt}|K_{it}; \theta_i\right) = \frac{\exp\left(\bar{w}_{it}\left(q_{ikt}, q_{ilt}\right)/\kappa\right)}{\sum_{\left(q_k^{'}, q_l^{'}\right)\in Q^2} \exp\left(\bar{w}_{it}\left(q_k^{'}, q_l^{'}\right)/\kappa\right)},$$

Where θ_i denotes all relevant parameters. Note that the possible choice set combinations $Q^2 = \{0,1,...,\bar{q}\} \times \{0,1,...,\bar{q}\}$ includes buying *nothing*, or buying from only one product.[12]

3.5.3.2 First stage decision

From the second stage, the inclusive value – expected maximum utility – given the first stage consideration set decision, can be obtained as

$$\bar{v}_{it}\left(K_{it}, \mathbf{A}_{it}\right) = \Gamma + \kappa \cdot \log\left(\sum_{\left(q_k^{'}, q_l^{'}\right)\in Q^2} \exp\left(\frac{\bar{w}_{it}\left(q_k^{'}, q_l^{'}\right) - F\left(K_{it}, \mathbf{A}_{it}\right)}{\kappa}\right)\right),$$

where Γ is the Euler constant. Then, the first stage decision maximizes the indirect utility

$$v_{it}(K) = \bar{v}_{it}(K) + \varepsilon_{iKt},$$

[12]For singleton consideration sets, the quantity support reduces to Q.

which yields the choice probability

$$\Pr\left(K_{it},\mathbf{A}_{it};\theta_i\right)=\frac{\exp\left(\bar{v}_{it}\left(K_{it},\mathbf{A}_{it}\right)\right)}{\sum_{K'\in\mathscr{K}}\exp\left(\bar{v}_{it}\left(K',\mathbf{A}_{it}\right)\right)},$$

where $\mathscr{K}$ is the set of all possible consideration sets (up to the size limit of 2) – including $\emptyset$.

3.5.4 Construction of the likelihood function

3.5.4.1 Matching the observed choice probability

We have characterized the choice probability of consideration set K, and the probability of purchase given K. To match the data, note that what is observed are the choice probabilities of a specific quantity combination, or, the empirical counterparts of

$$\Pr\left(q_{ikt},q_{ilt};\theta_i\right)=\sum_{K'\supset\{k,l\}}\Pr\left(q_{ikt},q_{ilt}|K^{'};\theta_i\right)\cdot\Pr\left(K^{'},\mathbf{A}_{it};\theta_i\right).$$

3.5.4.2 Likelihood with random coefficients

Given that each time series of choices by one individual is generated under one realization of random coefficients, we can write the likelihood of *all* the individual-trips, as

$$\mathscr{L}\left(\theta\right)=\prod_i\left(\int_{\theta_i}\left(\prod_t\Pr\left(q_{ikt},q_{ilt};\theta_i\right)\right)dG\left(\theta_i;\theta\right)\right),$$

Where $\left(q_{ikt},q_{ilt}\right)$ are observed quantities. The solver then minimizes $-\log\left(\mathscr{L}\left(\theta\right)\right)$ with respect to parameter θ.

3.5.4.3 Simulated maximum likelihood

The integral on θ_i is computed by simulation. To implement the simulated maximum likelihood method, we first take M draws of random coefficients *shocks* on β_{ih} and f_{ij},[13] denoted

[13] where β_{ih} is the h'th dimension of $\mathbf{x}_j$.

$\hat{\beta}_{mh}$ and $\hat{f}_{mj}$ for draw m, each independently from $\mathscr{N}(0, I_{J+H})$. Then, given parameters $\bar{\beta}_h$ and $\bar{f}_j$ and σ_β and σ_F, the individual i's random coefficient in draw m are determined by $\beta_{ihm} = \bar{\beta}_h + \sigma_\beta \cdot \hat{\beta}_{ihm}$, $f_{ijm} = \bar{f}_j + \sigma_F \cdot \hat{f}_{ijm}$; hence the random consumption coefficients on the product level are

$$\tilde{\beta}_{ijm} = \sum_{h=1,\ldots,H} \left(\bar{\beta}_h + \sigma_\beta \cdot \hat{\beta}_{ihm} \right) \cdot \mathbf{1}(x_j = 1)$$

We restrict γ and λ to be homogeneous across individuals.

We then maximize the likelihood function with respect to $\bar{\beta}_{mj}$, $\bar{F}_{mj}$, σ_β and σ_F, and other parameters, taking the draws as given. For each parameter value, the empirical counterpart of the likelihood function is given by

$$\mathscr{L}_N(\theta) = \prod_{i=1}^{N} \left(\frac{1}{M} \sum_{m=1}^{M} \left(\prod_t l_{it}(q_{ikt}, q_{ilt}; \theta_{im}) \right) \right).$$

3.5.5 Sub-sample

3.5.5.1 Choice of the sub-sample

To restrict computation burden at a reasonable level, we implement the structural model on a random sub-sample of 10% of individuals in the data (854 households), over all their in-sample trips. Because of dimensionality concerns, we focus on the 4 products that generate the highest overall sales (which consist of 30% of the total in-sample sales), and treat the rest as outside options. Finally, as previously indicated, we only consider consideration sets of size 0, 1 or 2.

The top 4 products are, respectively, Dannon Light N' Fit, Yoplait Original, Colombo Light and Yoplait Light. In the structural model, since we need to make across-trip comparisons of utility coefficients, we re-defined units to multiples of globally-available minimum units, rather than the minimum available units we use in reduced form. We adjust for choice probabilities for the unavailable units combinations.

Table 3.4: Distribution of number of products in the sub-sample

# products	frequency	share
0	10,476	0.63
1	5,726	0.35
2	359	0.02
3	17	0.00
4	1	0.00

Note: The table presents distribution of number of different purchased products in the same trip, in the sub-sample. Note that 0.1% of the sample purchased more than two different products.

3.5.5.2 Choice of characteristics

Given the choice of product set, we choose to focus on four characteristics – 3 brand indicators and the indicator for characteristics "light". These are denoted respectively, by $h = 1, ..., 4$.

3.5.5.3 Distribution of number of products

In the sub-sample, the distribution of the number of products chosen is shown in Table 3.4. Only 0.1% of the sub-sample purchased more than 2 different products, which justifies limiting the cap of the consideration set at 2. As a side note, zero products indicates purchase of another yogurt not in the set of interest.

3.5.5.4 Distribution of purchase quantity

Figure 3.8 summarizes distribution of choice-sets, and Figure 3.9 summarizes conditional quantity distribution given a singleton choice set. Multiple-product choice occasions are uncommon but present, which justifies allowing for only a parsimonious interaction term in the consumption utility specification.

Conditional on choice, quantity distributions are heavily skewed to the right – often with the mode quantity larger than 1. This can be rationalized by the scale economy generated from a per-variety consideration cost.

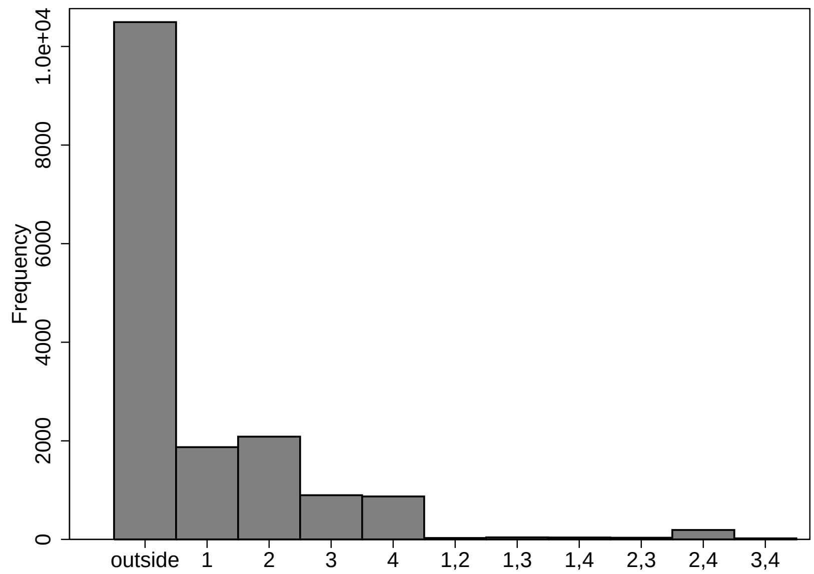


Figure 3.8: Quantity distribution of the sub-sample

Notes: This figure shows distribution of choices of bundles. This is creates using the full sample of households, on the choices of (combinations of) the top 4 products.

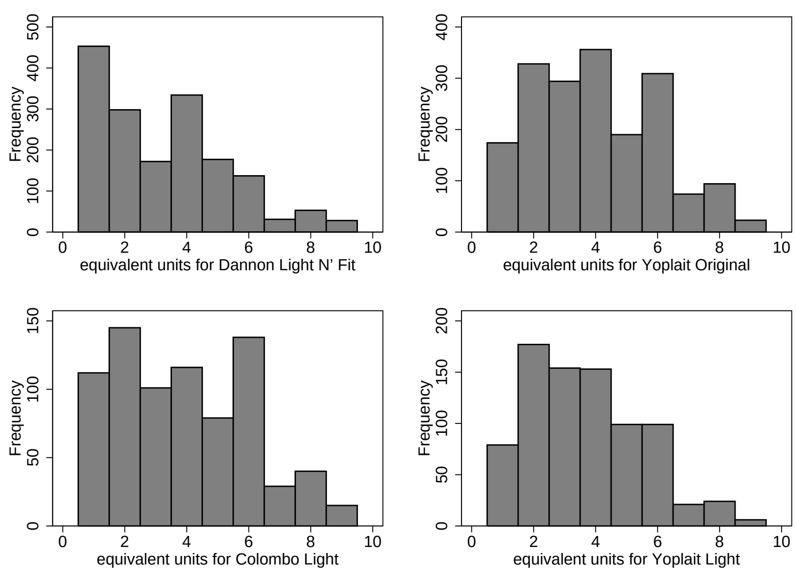


Figure 3.9: Quantity distribution of the sub-sample

Notes: The four figures depict quantity distributions given singleton bundle choice. As shown by Figure 3.8, multiple product choice is uncommon.

3.6 Parameter estimates

Table 3.5 reports all parameter estimates. $\bar{\beta}_h^\prime s$ capture the marginal *consumption* utility characteristics h, which is in turn multiplied by the log-transformed purchase quantity. $\bar{f}_j^\prime s$ capture the benchmark consideration cost.

The mean of the transformed $\tilde{\beta}_{ij}$'s are, 1.31 for Dannon Light, 2.75 for Yoplait Original, 2.78 for Colombo Light, and 1.96 for Yoplait Light. By defining random coefficients on characteristics, we capture the within-consumer correlation in demand, so that, for example, consumers who like Yoplait products will have higher choice probabilities on both Yoplait Original and Yoplait Light.

While using log-transformed quantity restricts the curvature of consumption utility for a single product, we do estimate the decreasing marginal utility across products – as captured by γ. For example, the estimates on $\bar{\beta}_h$ and γ imply that, for an *average* consumer, consuming 1 unit of Yoplait Original brings a utility of \$0.57,[14] and consuming the second unit further increases her utility to \$0.91. If – rather than buying the second unit of Yoplait Original – she instead consumes a unit of Colombo Light together with the first unit of Yoplait, her utility would have increased to \$1.08.[15]

On the other hand, the money metric for *mean* consideration cost, $\bar{f}_j/\lambda$, for all four products are, respectively, \$3.21, \$4.86, \$5.20 and \$4.73. Compared to the observed expenditure on each product,[16] consideration cost is worth 1.2 - 1.9 times the per-trip expenditure.

Estimate of σ_f^2 implies some variation of per-product consideration cost across individuals, but the 2.5-standard deviation (99th percentile) is away from zero. The large f_{ij}'s suggest that thinking about each product is costly, which in turn generates scale economy in the quantity choice. This is core to the discussion in Section 3.4.3. As a result, the consumer is incentivized to stay with fewer varieties and instead purchase large quantities. When purchasing multiple products, the total consideration cost is reduced by $\Delta f_2/\lambda = \$0.61$, suggesting a slight spillover in attention across products. Finally, when on feature advertising,

[14] $\tilde{\beta}_2 \cdot \log(1+1)/1.44$

[15] $\left(\tilde{\beta}_2 \cdot \log(2) + \tilde{\beta}_3 \cdot \log(2) + \gamma \cdot \log(2) \cdot \log(2)\right)/1.44$

[16] Conditional on product choice, the observed *average* per-trip expenditure on the 4 products are, respectively, \$2.61, \$2.92, \$3.49, \$2.46. The ratio between consideration cost and expenditure is 1.22, 1.66, 1.49 and 1.92.

Table 3.5: Parameters estimates

	par. est.	std. err.
brand coefficient for Dannon ($\bar{\beta}_1$)	2.10	0.16
brand coefficient for Yoplait ($\bar{\beta}_2$)	2.75	0.13
brand coefficient for Colombo ($\bar{\beta}_3$)	3.57	0.21
characteristics coef. for Light ($\bar{\beta}_4$)	-0.79	0.13
interaction of consumption utility (γ)	-1.16	0.13
price coefficient (λ)	1.44	0.06
consideration cost for Dannon Light N Fit ($\bar{f}_1$)	4.62	0.18
consideration cost for Yoplait Original ($\bar{f}_2$)	7.00	0.22
consideration cost for Colombo Light ($\bar{f}_3$)	7.49	0.24
consideration cost for Yoplait Light ($\bar{f}_4$)	6.82	0.23
changes in consid. cost for two products (Δf_2)	-0.88	0.20
changes in consid. cost under feature (Δf_A)	-0.60	0.09
scale of utility shock (κ)	1.62	0.10
variance of brand coef. ($\sigma^2_{\beta,1}$)	0.66	0.03
variance of light coef. ($\sigma^2_{\beta,4}$)	0.66	0.03
variance of mean consid. cost (σ^2_f)	1.74	0.05

Note: Estimates for the all parameters. Standard errors are asymptotic (numerical).

a product's consideration cost is reduced by $\Delta f_A/\lambda = \$0.42$.

3.7 Price and consideration-cost elasticities

3.7.1 Overall price elasticities

We compute the implied elasticities when one of the three products reduce prices by 5%. To do so, we first compute quantity choices based on each of 50 draws in the random coefficient, and then average them across draws. We then compute elasticities based on the percentage changes in the *average* quantity. We then decompose elasticities into changes in the consideration sets, and changes in quantity conditional on the consideration set.

Table 3.6 reports elasticities of the overall purchase quantities as response to a price change. The signs of the own- and cross- price elasticities are conventional, and the magnitude intuitive. For example, the own price elasticities for Danone Light N' Fit implies a 1.45% quantity increase, for each 1% price drop from the product.

Table 3.6: Implied elasticities

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-1.45	0.10	0.24	0.12
Yoplait Original	0.16	-2.49	0.02	0.32
Colombo Light	0.21	0.06	-1.89	0.24
Yoplait Light	0.06	0.13	0.18	-1.88

Note: The i, j element is the elasticity of total purchase quantity of product j, on price change of product i, or $\frac{\partial Q_j}{\partial P_i} \cdot \frac{P_i}{Q_j}$.

Note that the numbers here are conditional on within-category purchases. Bell et al. (1999) find that within-category switching accounts for 75% of the total price elasticities, which hints that the within-category elasticity we find is a major part of the total elasticities.

3.7.2 Decomposition of price elasticities

With a model of consideration set formation and purchase choice, we can decompose the overall price elasticities into two parts. First, a reduction in price drives up quantity choice conditional on consideration; second, it also facilitates consideration set membership, in the sense that a consumer might now find it attractive to consider a product.

The decomposed elasticities are presented in Table 3.7 and 3.8. We find that consideration set membership is very responsive to price changes, as the own-price elasticities shown in Table 3.7 are close in magnitude compared to the overall elasticity estimates. This is not surprising, given the nature of scale economy driven by the model estimates. Because considering a product is costly apart from price, the consumer will be likely to find it optimal to ignore an entire product when the prices are high.

On the contrary, elasticities given consideration is low – when a consumer is only considering a single product. Here, products are local monopolists in singleton sets, where the consumer can only choose a single product or the outside option.

Table 3.7: Implied consideration-set formation elasticities

	(A)	(B)	(C)	(D)	others
Dannon Light N Fit	-0.98	0.08	0.18	0.10	0.20
Yoplait Original	0.08	-1.86	0.06	0.28	0.55
Colombo Light	0.15	0.04	-1.27	0.25	0.15
Yoplait Light	0.06	0.16	0.14	-1.37	0.09

Note: The i, j element is the elasticity of consideration probability of product j, on price change of product i, or $\frac{\partial \Pr(j \in K_{it})}{\partial P_i} \cdot \frac{P_i}{\Pr(j \in K_{it})}$.

Table 3.8: Implied elasticities given consideration

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-0.42	0.00	0.00	0.00
Yoplait Original	0.00	-0.73	0.00	0.00
Colombo Light	0.00	0.00	-0.84	0.00
Yoplait Light	0.00	0.00	0.00	-0.87

Note: The i, j element is the elasticity of purchase quantity of product j, given that j is the only product considered, on price change of product i; or: $\frac{\partial Q_j}{\partial P_i} \cdot \frac{P_i}{Q_j}$ given that $K_{it} = \{j\}$.

3.7.3 Consideration cost elasticities and its decomposition

We then compute the elasticities of consumer purchase decisions to a change in consideration cost. To do so, we simulate choices when a product is on a hypothetical feature advertisement that reduces consumer consideration cost by 5%. Table 3.9 reports overall consideration cost elasticities, while Table 3.10 and **??** decompose them into effects on consideration set and on conditional quantity choice, similar to the previous section.

It is worth noting that the overall consideration cost elasticities imply that a 1% consideration cost reduction converts to 1.70-3.75% increase in overall purchase quantities.

Table 3.9: Implied consideration cost elasticities

	(A)	(B)	(C)	(D)
Dannon Light N Fit	-1.70	0.19	0.44	0.30
Yoplait Original	0.32	-3.66	0.25	0.67
Colombo Light	0.26	0.12	-2.39	0.30
Yoplait Light	0.14	0.40	0.27	-3.75

Note: The i, j element is the elasticity of total purchase quantity of product j, on change of search cost of product i, or $\frac{\partial Q_j}{\partial F_i} \cdot \frac{F_i}{Q_j}$.

Table 3.10: Implied consideration cost elasticities on consideration set formation

	(A)	(B)	(C)	(D)	others
Dannon Light N Fit	-1.88	0.03	0.26	0.14	0.34
Yoplait Original	0.15	-3.76	0.23	0.56	0.79
Colombo Light	0.20	0.10	-2.62	0.24	0.25
Yoplait Light	0.13	0.24	0.30	-3.98	0.24

Note: The i, j element is the elasticity of consideration probability of product j, on change of search cost of product i, or $\frac{\partial \Pr(j \in K_{it})}{\partial F_i} \cdot \frac{F_i}{\Pr(j \in K_{it})}$.

Multiplied by the percentage decrease in consideration cost from feature advertising (12-16%), this means that featuring a product is associated with 27-64% sales increase.[17] Of course, given that we condition the analysis on category purchase, this calculation does not take into account the effect of feature in driving a consumer to the refrigerated section.

3.8 Feature advertising and price discounts

3.8.1 Informative versus persuasive feature: intuition

When feature advertising is complementary to consumption (Becker and Murphy, 1993) – as commonly imposed in the empirical choice modeling literature (Chintagunta, 1993; among others) – feature advertising increases the relative importance of consumption utility, and comparatively, downplays the role of price.[18] Hence, featuring a products decreases its price elasticities, and this strategy should be seen as a substitute to a price discount.

[17] The highest sales increase is on Yoplait Original.

[18] This property comes from the logit structure. To see this, specify a logit model

$$u_1(f,p) + \varepsilon = \beta_1 f - \beta_2 p + \varepsilon_1,$$

and

$$u_0 = \varepsilon_0.$$

Because market share is

$$s(f,p) = \frac{\exp(\beta_1 f - \beta_2 p)}{1 + \exp(\beta_1 f - \beta_2 p)},$$

it is apparent that

$$\begin{aligned} \mathscr{E} &= \frac{\partial s}{\partial p} \cdot \frac{p}{s} \\ &= -\beta_2 p (1 - s) \end{aligned}$$

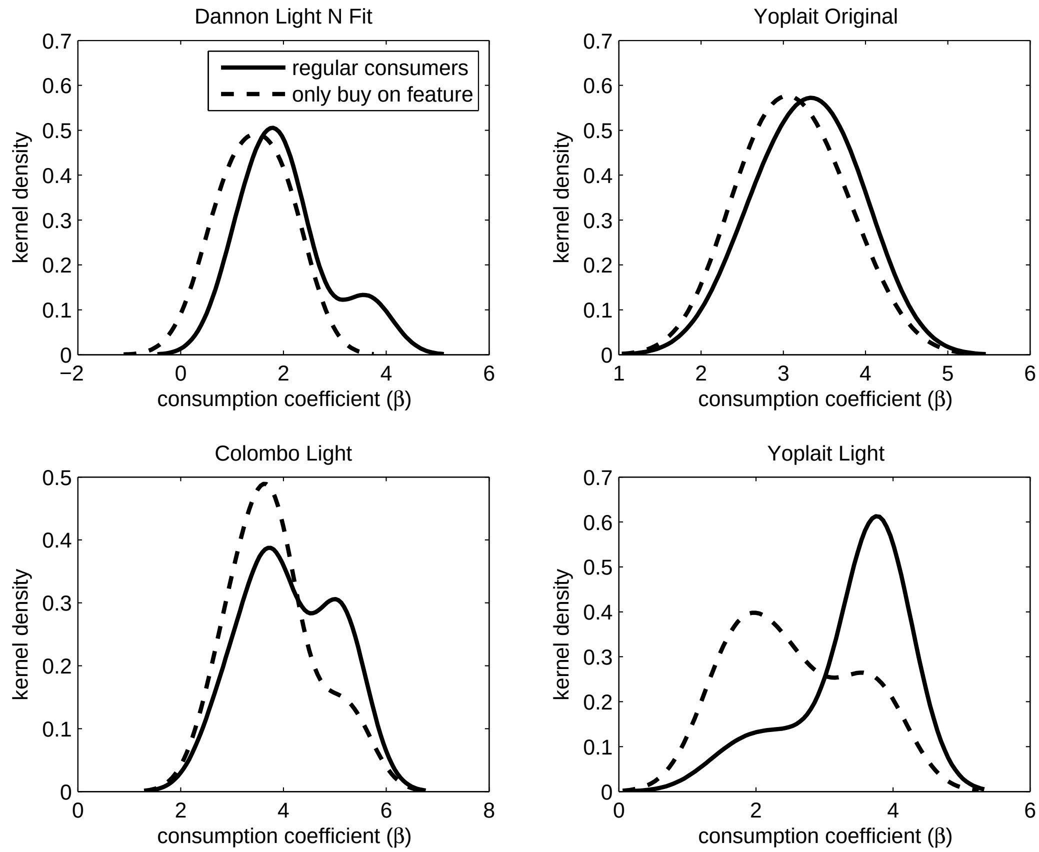


Figure 3.10: Distribution of utility coefficients for different consumers

Notes: These figures present the posterior distribution of consumption utility coefficients $\tilde{\beta}_{ij}$ across consumers, for different products. The solid lines plot the posterior (Kernel) density for consumers who purchases the product without it being featured – i.e "regular consumers". On the other hand, the dashed lines plot the density of $\tilde{\beta}_{ij}$ for consumers who purchase *only when* the product is on feature. The bandwidth for Kernel estimators is set to 1.

This implies,[19] that feature and price discounts should be asynchronous, because consumers facing feature are less price elastic. This is at odds with the observation that they are usually seen simultaneously (Figure 3.2).

In this framework, however, feature advertising is seen as a substitute to discounts in the consideration stage, following the previous argument. However, it is seen as a complement to price discounts in the purchase stage. To intuitively see the second point, Figure 3.10 shows posterior distribution of consumption utility coefficients ($\tilde{\beta}_{ij}$), separately for two

and $|\mathscr{E}|$ should be decreasing in s, which is an increasing function of f.

[19]If we do not consider the role of feature advertising on providing price information. In this context, this is abstracted away given that we focus on choices conditional on category purchase.

Table 3.11: Changes in overall price elasticities when on feature

	(A)	(B)	(C)	(D)
Dannon Light N Fit	0.91	1.17	0.56	0.84
Yoplait Original	0.90	0.87	3.80	0.96
Colombo Light	0.66	1.08	0.94	0.85
Yoplait Light	1.14	1.04	1.22	0.98

Note: Each cell i, j presents the ratio between i's *counterfactual* elasticity to j when j is on feature with probability 1, to the benchmark elasticity presented in Table 3.6 (when j is featured as data shows). That is, the i, j element is $\frac{\mathscr{E}_{ij}|_{f_j=1}}{\mathscr{E}_{ij}|_{f_j}}$.

groups of consumers: one group who regularly purchases the product when it is not on feature, and another group who only makes the purchase when the product is featured. The second group of consumers are the ones with lower consumption utility, or in other words more price elastic. Hence, featuring a product raises the overall quantity elasticities, which then implies that prices should also go down to "further persuade" a consumer into purchasing. This is consistent with the empirical regularity that featuring a product is associated with a price discount.

3.8.2 Feature advertising and price elasticities

With a model where feature reduces consideration cost (and hence is informative), we can then investigate the effect of featuring a product on its own- and cross- price elasticities. Table 3.11 reports the ratio between the (overall) price elasticities when the row product is on feature with probability 1, and the benchmark price elasticities when products are on/off feature as observed. We find that being on feature *reduces* the own-price elasticities by a magnitude of 7%-9% – which at a first glance suggest that feature advertising is a substitute to price discounts.

On the other hand, decomposing the elasticities, we find that quantity elasticity conditional on consideration set membership has increased as a result of being on feature, presented in Table 3.12.[20] This is because, when consideration costs are reduced by featuring, some consumers with lower $\tilde{\beta}_{ij}$'s are going to start considering the product – these con-

[20] In this table, since we focus on quantity elasticities given singleton consideration sets, cross elasticities are by construction zero and are ignored.

Table 3.12: Changes in conditional quantity elasticities when on feature

	ratio between own-elasticities
Dannon Light N Fit	1.08
Yoplait Original	0.96
Colombo Light	1.02
Yoplait Light	1.12

Note: See note in Table 3.11, except that this table reports ratios in quantity choice elasticities conditional on consideration.

sumers would otherwise think it is unfruitful to think about product j due to the low $\tilde{\beta}_{ij}$'s. At the same time, the low $\tilde{\beta}_{ij}$'s suggest that these are the higher-elasticity consumers, whose participation then drives up the overall price response.

3.9 Concluding remarks

This paper quantifies a consumer's cost of processing product information that are less salient, such as to understand the nutrition content before buying a yogurt. Such consideration costs are fixed to quantity, therefore generate scale economy in a consumer's quantity choice. Therefore, a consumer is discouraged from spreading her consideration effort, and purchase decisions, over a vast number of products (even when she loves variety). However, if variations in salient product characteristics (such as price) alters consideration set membership, the newly included products display discontinuous jumps in the purchase quantity.

We utilize this identifying property, and quantify the magnitude of consumer consideration cost. Specifically, we develop testable implications by looking at consumer purchase quantity at her maximum accepted price, and find that quantity jump is far larger than a model without costly consideration can justify. We then build a structural model of multiple product choice and the subsequent quantity choices, with an endogenous costly consideration decision in the first stage. Notably, conventional discrete-continuous models of multiple product choice do not accommodate the discontinuity in quantity choice, which is our key identifying pattern. Consequently, our model deviates from canonical models, at the cost of additional, manageable computation burden. As by-products, the model also allows for flexible nonlinear price structures and discrete quantities.

We estimate the model using IRI behavioral scan panel data, in the yogurt category. In our case, the average (monetized) consumer consideration cost per yogurt product is between \$3.2 and \$5.2 in a given trip. This is about 1.2-1.9 times the total expenditure on each product per trip. The magnitude of consideration cost indicates that prices are not the dominant explanation of which subset of product to choose from; and conversely, this, explains why price elasticities are so small on quantity choice conditional on purchase. In fact, we find that elasticities of quantity given purchase is about 1/3 of the overall price elasticities, implying that the main effect of price promotion is to penetrate the consideration barrier of consumers – to incentivize them to purchase (as opposed to purchasing more).

Finally, we find that feature advertising reduces the consideration cost by \$0.42, which is on the order of 10% of the total consideration cost. This drives more consumers to purchase, but because of preference heterogeneity, mostly affects consumers with lower tastes – those consumers would otherwise not purchase. This creates a selection problem, that consumers who are attracted during a feature promotion are the ones with lower tastes and higher price elasticities. Hence, our model justifies the synchronized use of price and feature promotions.

Chapter 4

Inconvenience versus Risk in Consumer Channel Choice

4.1 Introduction

Online retailing plays an increasingly important role. According to US Census Bureau, the share of retail trade conducted online has increased from 0.19% in 1998, to 5.22% in 2012.[1] Even for experienced goods such as wine or fashion, online retailing is becoming increasingly common. For instance, in data covering the Dutch retail apparel industry, we find that the share of consumer expenditure spent online has risen from 5.5% in 2007, to 12.5% in 2014. The increasing popularity of the online channel changes the strategy of traditional sellers, and many of them are expanding into online retailing, adjusting their store locations accordingly.

We are interested in understanding a consumer's channel choice within a retail chain. A more comprehensive account of this decision is of specific interest to managers who consider expanding online versus establishing more traditional retail stores, and needs to know how large the cannibalization effect is from either action. It is also of interest to policy makers who seek to understand pricing and location choices as outcome of substitution

[1] U.S. Census Bureau, Annual Retail Trade Survey, 2012.

between retailers and between channels of each retailer, and design tax or product return regulation policies.

Specifically, we seek to understand *how* and *why* consumers substitute between the online and traditional (in-store) retail channels. In the literature, substitution between channels is often confounded with substitution across different retailers, who might be offering different products or services. For example, Goolsbee (2000) and Ellison and Ellison (2009) study the effect of sales tax on online-retailer demand, and Forman et al. (2006) study variation in distance on the market share of online booksellers, but none of them can isolate substitution between channels from substitution between retailers. This makes the result less applicable to a manager who is trying to separate competition effect and own cannibalization effect, or to a policy maker who is assessing merger and acquisition cases that involve different chains in multiple channels.

In this paper, we study the sensitivity to travel cost – as a main component in her shopping cost – in consumer choice among different channels, provided by the *same* retail chain. Specifically, we exploit exogenous variation in a *single* consumer's distance to the closest outlet of a *given* chain, which can be considered as variation in the travel cost for shopping in-store. This variation comes from consumer house-moving and from outlet entry and exit. We find that the consumer substitutes away from the physical store channel when living further away from outlets of that chain. In addition, we measure the consumer's sensitivity to distance over time, and find that distance has an increasing effect on channel choice. We provide descriptive evidence that the sensitivity to distance in a consumer's channel choice is 3 times as high in 2014 than in 2007, and it is 3 times as high when a consumer has shopped 10 times online in the past, than when she had no or little experience online. This indicates that online and off-line channels are substitutes, to an increasing degree in calendar time and/or consumer experience.

We further pursue the second question: *why* are the two channels increasing substitutes? Among prior explanations, Prince (2007) focused on changes in distributions of consumers and retailers heterogeneity, but these explanations cannot apply to our analysis which keeps consumer / retail chain combinations constant. We propose an alternative explanation, highlighting the changes in a consumer's perceived risk, interacted with her inability to sample a product prior to purchase online (the latter mechanism exists in prior literature, e.g. in Ofek et al., 2011). Specifically, we propose that a consumer's perceived probability of receiving an inferior product (hence the term "perceived risk") decreases over time.[2] Because a consumer shopping in the traditional channel can verify product quality before the purchase, off-line shopping naturally serves as an "insurance" against the risk of inferior product, and the travel cost associated with off-line shopping can be viewed as the insurance premium. When the perceived risk goes down, consumers are less reliant on verifying product quality before purchase; hence, their choices are more sensitive to variations in travel cost. This is an explanation to the rise of online channel market shares, and it generates a hypothesis that exogenous variations in expenditure leads to decreases in online purchase probability, and the magnitude of this effect is smaller in later calendar time periods, or for a consumer with more online-shopping experience.

To test this hypothesis, we seek to overcome a reverse causality problem, that consumers who decide to go online are more cautious due to the the risk, and consequently spend less. We attempt to solve this problem by exploiting exogenous variation in average price, consumer income and season-specific products that she purchases. Specifically, we assume that a consumer's income and a chain's overall price level are pre-determined, and are exogenous to her idiosyncratic shocks in channel choice. We also assume that specific seasonal needs, such as buying T-shirts in summer and coats in winter, are orthogonal to unobserved shocks in channel choice. Of course, confounding explanations do exist: consumers with higher income might systematically be less averse to online risk; chains with higher price might be of better quality assurance online; and consumers might be driven to purchase online in winter because of higher, weather-induced travel cost. We address these concerns by controlling for a rich set of individual, chain, time and weather fixed effects.

Using purchase data for a panel of Dutch consumers of retail apparels from 2007 to 2014, we find that trips with lower expenditure – due to reductions in income, purchases of summer clothes, or shopping during discount seasons – are associated with higher online-shopping tendency. In addition, the sensitivity of channel choice to exogenous variations

[2]This remains true even when the consumers can return products because returns delay the timing of consumption, and the dis-utility for delayed consumption reasonably goes up with expenditure.

of expenditure decreases over experience and time. This shows that expenditure acts as a "price" to online shopping, and this price is higher in early years and when a consumer is less experienced.

In light of the reduced-form evidence, we then structurally characterize a consumer's choice whether to shop online and her expenditure given the choice, by constructing a simple model while controlling for a rich set of individual, chain and time fixed effects. Specifically, we construct a simple model of consumer choices, which numerically solves for average online-shopping tendency and expenditure, as functions of calendar time, a consumer's past online shopping experience, and her distance to the nearest store outlet. Meanwhile, we can non-parametrically characterize the average choice and expenditure profiles, while controlling for individual, store and time fixed effects. These empirical moments can then be directly matched to the model-predictions, in order to determine the structural parameters. In this way, we isolate the mechanism of interest – which we model structurally – from non-central variations which are controlled, but kept in reduced form.

Among the two factors, experience and time, we find that past online-shopping experience has the stronger effect. After the first three online trip, a consumer would find online shopping much safer, in terms of a 20%-lower perceived risk, than what she believed when she had no experience. Further experience decreases the perceived risk by another 10-15%. Compared to the experience effect, evolution of the market environment (for instance, better websites or return services) during 2007-2014 decreases the perceived risk by 12%. This indicates that changes within a consumer is the main explanation of why the online stores of given retailers slowly gain more market shares. In fact, we find that a very experienced consumer would be 3 times as likely to shop online, as when she is completely inexperienced; conversely, the extra experience would then save her 1.27 kilometers of travel cost on average, every time she needs to buy apparels.

Our work contributes to the literature in two ways. First, to the best of our knowledge, we provide the first set of estimates on channel choice sensitivity to distance, keeping consumer and retailer identities fixed. Prior literature usually confound channel choice sensitivity with store choice sensitivity (Prince 2007; Forman et al. 2006; among others), and hence downward biases the estimates of distance effect. The small sensitivity to distance *across* retailers gives the wrong impression that demand is also insensitive to distance *within* a retailer, which then causes a manager to underestimate the within-chain cannibalization effect. In fact, we find that for an experienced consumer living at the average distance, opening a new outlet at her doorstep would take away 10% sales from the online store. Conversely, if there were no online stores, opening one will cannibalize part of the off-line sales, apart from any potential brand-expansion effects.[3] In addition, since channel substitutability increases with consumer experience and calendar time, so is the cannibalization effect. This puts caution to a traditional retailer considering opening an online branch.

Second, we contribute to the understanding of why retail channels are increasingly substitutable. We offer a new explanation, that increasing substitutability between retail channels is caused by changes in online-shopping risk (Ofek et al., 2011) over time and experience. More specifically, we develop a model that emphasizes a consumer's ability to examine a product before purchase; then, despite costly in travels, shopping off-line provides an insurance against the perceived risk of receiving an inferior product. Both our reduced-form tests and structural parameter estimates indicate that the evolution of consumer experience is an important driver of the reduction in risk. This then implies that the majority of online-shopping risk is in fact *subjective*, and an inexperienced consumer overly estimates her potential risk of not getting what she intended to. Therefore, she is overly cautious both in her decision of going online, and in her expenditure if shopping online. Despite that such expectation bias can be corrected over many shopping trips, these tend to occur over a long period. Therefore, potentially, a retailer might consider policies that changes consumer belief; one example of such would be penetration pricing, exclusively on the online shop.

The rest of the paper is organized as follows. Section 4.2 reviews the literature. Section 4.3 presents the data and summary statistics, and discusses patterns of channel choice's sensitivity to distance over time and consumer experience. Section 4.4 presents a reduced-form empirical model, and discusses identification strategy, in order to test for the role of online risk (and its reduction) in changes in consumer channel choice patterns. In light of the empirical evidence, Section 4.5 presents the model and estimation strategy. Section 4.6 discusses estimation results, model fit and managerial implications. Finally, Section 4.7

[3]In our empirical exercise, we do not quantify the expansion of the total sales of a given chain.

concludes.

4.2 Related literature

Over time, consumers have been steadily migrating purchasing to online channels (Ansari et al., 2008). Online retailing offers an appealing transaction channel (Peterson et al., 1997). However it may be an inferior alternative to offline retailing with respect to evaluating product quality or match value, especially in the case of experience goods. In this paper, we study the case of consumer purchases for apparel and we are in particular interested in the motives of consumers to buy apparel online.

Buying online may be motivated by travel cost or distance between the consumer and a store. Thomadsen (2007) and Chintagunta et al. (2012) report high estimates of travel costs in a purchasing context. Even for high price durable goods such as cars, demand is often surprisingly local (Albuquerque and Bronnenberg, 2012; Bucklin et al., 2008) suggesting also that travel costs are high. Online channels allow consumers to greatly reduce travel costs, or eliminate them altogether; hence, one might suspect high sensitivity in consumer choices between the two channels.

However, Prince (2007) documents that this is not the case in early years – demand elasticity to prices among online and traditional retailers became significantly different from zero only until 1998. Similarly, Deleersnyder et al. (2002) find that for newspapers, the establishment of an online distribution channel has no immediate effect on circulation and advertising revenue from the traditional channel. Finally, Forman et al. (2006) document that the presence of a traditional retailer in the vicinity changes the market share of a particular product online.

We add to this literature by isolating the sensitivity to distance in channel choice, conditional on the same consumers and retailers. We document that the distance between consumers and the nearest chain outlet is informative about the propensity to buy online, and the sensitivity to distance is increasing over time. In addition, since the two channels offered by the same retailer are less differentiated than across different retailers, it is natural to understand that we find much higher sensitivities when conditional on the retailer choice.

This then speaks to the practitioner discussion on size of cannibalization effects between channels.

On the other hand, online purchasing, while convenient, may pose a risk to some consumers. In a study of channel migration in the grocery trade, Bell and Song (2007) find that consumers are quicker to adopt services from a national online grocer when they live in an area that already contains many adopters. Choi et al. (2010) find that this effect is especially strong during the early presence of the online channel, suggesting that the need for consumers to see evidence of the concrete benefits of the online channel is highest when no information about it is present. Huang et al. (2009) find that consumers do more online processing of information and rely more on reviews for experience than for search goods. Ofek et al. (2011) also suggest that buying online is risky, and relate the source of risk to an online-shopper's inability to sample the product prior to purchase. Our explanation to why the online and traditional channels are closer substitutes, is most closely related to this literature: that improvement of a seller's information provision and the authority's regulation on service quality decreases a consumer's perceived risk.

4.3 Data, sample selection and descriptive evidence

4.3.1 Purchase data

Our primary purchase data are from GfK's JURY panel in the Netherlands. The panel covers 861,550 purchases of fashion items for 19,291 households over the period 2007Q1-2014Q3 representing €26.9M. Purchases are recorded using a diary. The data are at the panelist-trip-item level. For each purchase, we observe quantity, expenditure, whether it was accompanied by a deal and whether the purchase was carried out online. Of all purchases made, 49% is recorded to be accompanied by a price deal. Figure 4.1 shows the expenditure share of online purchasing by year. What can be observed is that purchasing fashion online is gaining strongly in popularity.

The keys of the data have the following descriptor fields. For each panelist, we observe demographic characteristics such as education, income, age, employment status, etc. We

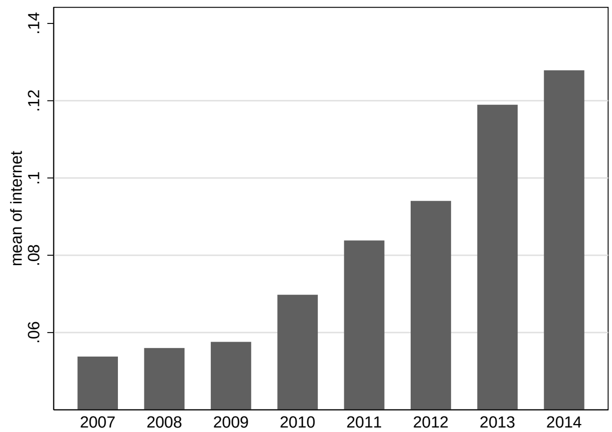


Figure 4.1: Expenditure share of online purchases in fashion

Table 4.1: Descriptive statistics: Panelists

variable	*N*	mean	standard deviation
days in panel	19291	998.578	986.331
number of recorded trips	18863	22.583	36.091
units	19291	53.701	93.709
euros	19291	1393.466	2637.493
euros/trip	18863	69.880	73.086
net monthly income	14720	2469.477	1010.495
recipient female	19291	0.591	0.401
recipient age	19291	40.366	19.136
retired	19291	0.145	0.340

also observe the location of each consumer at the postal code "plus" level: the highest postal code resolution is 4 digits plus 2 letters, e.g., 5037 AB, and we observe the location of consumers recorded as "5037 A." The median distance to the closest 5-digit postal code is 0.34 km. Table 4.1 lists the mean and standard deviation of selected purchasing activity and demographic characteristics for our panelists. Households stay in the panel for 3 years on average and record 23 trips with an average expenditure of €70 per trip. The net monthly income averages €2470. Of all purchases made, 60% is for female recipients, and the average recipient age is 40 years.

For each trip, we observe the time, the store chain (e.g., H&M, Zara), and the retail segment (e.g., department store, sports store). Table 4.2 lists expenditure and expenditure shares for the top 5 chains. As can be seen the retail concentration of fashion is not very high, with the joint share of the 5 largest chains remaining well under 20%.

Table 4.2: Descriptive statistics: Top chains

Chain	expenditure	expenditure share
C&A	1686837	6.27
Miss Etam	835403	3.11
H&M (Hennes & Mauritz)	831199	3.09
Vroom & Dreesmann	628170	2.34
Esprit	505692	1.88

Table 4.3: Descriptive statistics: Top products

Product	expenditure	expenditure share
Shirt	6035984	22.45
Pants-Trousers	2905253	10.81
Jeans	2855154	10.62
Jack-parka	1829248	6.80
Vest	1721081	6.40

Finally, for each item purchased, we observe brand, department (e.g., Jackets, Outerwear), and product (e.g., shirts, trousers). Table 4.3 lists the 5 largest product categories in our panel along with their expenditure shares. The top 5 product categories account for almost half of the fashion purchases.

4.3.2 Online purchase experience

Using the purchase record data, we construct a consumer's online purchase experience in a shopping trip. A consumer's experience at time t is defined as the number of *online* shopping trips done before t. Table 4.4 summarizes distribution of total number of online trips (individual level), number of online trips done in the past (trip level), and the duration in sample before the first online trip (individual level). Note that the third column, months in sample before the first (online) trip, only focuses on the subset of individuals who has at least shopped online once.

4.3.3 Official postal code coordinates data

We obtain the Dutch official mapping data for postal codes and coordinates, the "Geo Suite" database. The data covers the coordinates (latitude and longitude) for all 4, 5 and 6 digit postal code in the Netherlands. A 6-digit postal code covers on average 8 addresses, while a

Table 4.4: Descriptive statistics: Online trip history

	Total number of (online) trips	# trips prior to t	Months in sample before first trip
1%	0	0	0
10%	0	0	0
50%	0	0	5
90%	2	7	44
99%	31	36	84
# obs	14,989	201,626	3,916

4-digit postal code is on the accuracy level of a street.[4] Because we only have 5-digit postal code data from the household panel, we map the 5-digit postal code to coordinate file to obtain household locations, while mapping the 6-digit file to obtain outlet locations. This can be viewed as taking simple average assuming that a household always resides in the center of a given area. As mentioned in the description of purchase data, the 5-digit postal code provides precise information on location.

4.3.4 Retail outlet location data

We complement the micro panel data by the Orbis company information database, from Bureau van Dijk.[5] The database consists of all registered companies and their branches, with contact information (address and postal code), legal information (e.g. active or dissolved/bankrupt), and others. This database provides information retail outlet location as well as entry and exit. With our aim, we extract all companies in the retail and wholesale industries in the Netherlands, with their outlet locations, addresses and whether each outlet is active at each of the recorded dates. We then define “entry” as the first year and month when an outlet is registered as active in the Orbis data, and “exit” as and year and month when an outlet is registered as dissolved or bankrupt. This results in a monthly data of outlet entry and exit.

Since Aimark data does not contain detailed outlet location data, we assume that the consumer considers the distance between her home address and the closest outlet. To do so, we first calculate their great circle distance, i.e. the length of the arc between coordinates

[4] http://en.wikipedia.org/wiki/Postal_codes_in_the_Netherlands
[5] http://www.bvdinfo.com/

(x,y) of the household i, retail outlet o at time t:

$$D_{iot} = r \cdot \arccos\left(\sin x_i \sin x_o + \cos x_i \cos x_o \cdot \cos\left(y_i - y_o\right)\right)$$

Then, we find minimum distance among all active outlets of a retail chain j, that is $D_{ijt} = \min_{o \in j} D_{iot}$.

4.3.5 Daily weather data

Finally, we obtain daily weather data from the Royal Netherlands Meteorological Institute.[6] These data-sets consist of daily weather records from a total of 37 stations, each provided with its coordinates. The records include temperature, duration of sunshine, duration of precipitation, wind speed, and others. For each trip, we merge the weather records from the closest station to the trip location (that is, outlet location), on the trip date.

4.3.6 Sample selection

We merge GfK consumer panel data on consumer apparel with Orbis data on chain/outlet location and entry/exit time, as well as KNMI daily weather data. We only focus on the top 18 chains in total in-sample sales, after excluding all unidentified stores and smaller retail chains, plus one major online-only retailer. This takes away 60% of the data, yet this is mostly a result of missing chain identity. Since it is impossible to cleanly analyze within-retailer choices of channels, without knowing the exact chain identity, this sample selection step is crucial. We further select observations that have non-missing trip date, distance from home address to store address, as well as weather information on the given trip date. Finally, we pick trips that are no further than 30km, since very long trips are usually associated with other activities such as family events.[7] Altogether, These criteria give 259,327 observations. Table 4.5 summarizes our sample selection criteria.

[6] Koninklijk Nederlands Meteorologisch Instituut, KNMI

[7] Only 30% observations are within 30km distance, but this mainly reflect missing of the distance measure (e.g. due to unidentified chain ID).

Table 4.5: Sample selection

	percentage
chain ranked top 20	0.373
distance from home to outlet not missing	0.315
trip date not missing	0.368
weather on the trip not missing	0.987
closest outlet within 30km	0.306
all above criteria	0.301
obs.	861550

Notes: This table reports our sample selection criteria.

Lastly, we collapse these data onto household-chain-date level. A consumer on average buys 2 items in a given trip. This yield 126,936 observations for estimation.

4.3.7 Online shopping and expenditure over experience and time

We document the sensitivity of online shopping choice probability and expenditure to changes in distance, and the changes of this sensitivity over experience and time. Since distance (as a measure of travel cost) is clearly an important component in the total off-line shopping cost, changes in the sensitivity to shopping cost is informative of the changes in channel substitutability.

We measure experience as the number of *online* trips done in the past, as documented in Section 4.3.2. Controlling for this measure of experience, we separately regress a consumer's online shopping tendency to her distance to the nearest outlet, by half-year time intervals. Changes in the marginal effect of distance indicate changes in channel substitutability, over calendar time. We also control for individual-chain combined fixed effect, weather fixed effects, and net income. Next, we perform a similar analysis of online shopping on distance, but by trips with different shopping experience. In this case, we control for time interval fixed effect, along with individual and chain fixed effects, and observables.

As documented in the top panels of Figure 4.2, we find that the sensitivity to distance increases in later years, in particular during years 2013 and 2014. This indicates that online and traditional channels became better substitutes with time. On the other hand, consumers with more online shopping experience in the past tend to shop more frequently online –

and the effect seems much larger than that of calendar year. However, if consumers are heterogeneous in their preference towards online shopping, correlation in past shopping experience and current behavior might reflect such heterogeneity. Therefore, the reader should be aware of the possibility of selection, in the upper-right panel.

We also produce descriptive pattern of off-line expenditure to distance,[8] separately by time and experience. We find that overall, expenditure is positively correlated with distance. This might reflect that consumers who chose to go to the store *despite* larger distance might have more to buy in the first place. In later years, expenditure is more sensitive to distance due to that fewer consumers with small expenditure still chooses to go off-line. The same logic goes for the lower-right panel on expenditure and distance over experience, but we do not observe clear pattern.

Overall, this section shows that shopping trips done later (in terms of calendar years), or done when a consumer is more experienced with online shopping, are more sensitive to distance. If we interpret distance as a measure of travel cost – intuitively, this is a crucial part of the shopping cost difference between online and off-line channels – the results here then imply that the substitutability between the two channels increases with time and experience.

4.4 The response of online shopping to trip expenditure

4.4.1 Overview

The evidence presented in Figure 4.2 can be rationalized by changes in the perceived risk in online shopping. Specifically, if the consumer believes that when purchasing online, there is a risk of not receiving the item that she intended to buy – which might be the case if she receives an item of the wrong size or color, or an item of lower quality than advertised online – this might discourage her from shopping online especially when expenditure is high.

In fact, when the consumers hold such belief, a fraction of expenditure online would be

[8] We do not analyze online expenditure because online trips are relatively rare, and therefore, online expenditure contains more noise than the off-line counterpart.

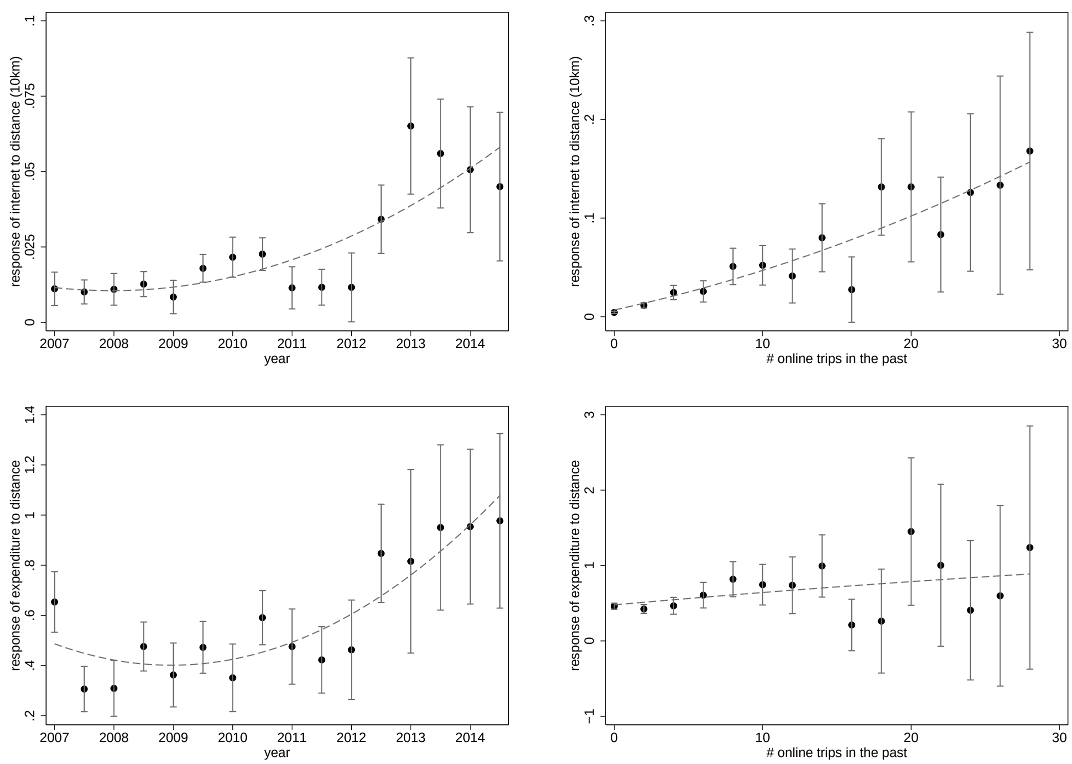


Figure 4.2: Response to distance in channel choice and expenditure

Note: This figure presents OLS estimates of a consumer's channel choice and expenditure, as response to changes in the distance to the closest outlet. In the upper and lower right panels, we estimate the linear model separately for each half-year time interval, controlling for the number of online trips in the past. In the upper and lower left panels, we do the opposite and control for half-year dummies, while estimating the model by trip history. In addition, we control for individual and chain fixed effects, as well as net income and weather fixed effect.

in exchange for nothing. Therefore, the *effective price* of purchasing online is proportional to expenditure, and the consumer who plan to spend more – assuming that expenditure is exogenously given regardless of preference – will tend to shop off-line. On the other hand, as experience accumulates which triggers learning, or as the quality of service from the supply side evolves with time, the perceived online risk will be reduced, thereby lowering the sensitivity of online shopping to changes in expenditure.

We test whether a consumer's tendency to shop online changes, if her expenditure in the trip is varied by price, income, and the type of product she buys. We argue that these variations are exogenous to a consumer's idiosyncratic preferences towards the online channel. We find that an increase in total expenditure in a shopping trip reduces the likelihood of purchasing online, but the effect size is smaller in later years in the data. We also find that, within the same (sub)set of consumers who would have had the same amount of experience in the end, online purchase's response to expenditure is smaller when they have accumulated more experience. These result support the hypothesis that reduction in online risk over experience and time is an important explanation to the rise of online shopping.

4.4.2 Reduced form specification

To test whether the sensitivity to expenditure varies with time, we estimate a reduced form model of a consumer's online shopping decision's response to her expenditure in the trip:

$$I_{ijt} = c + (\alpha_0 + \alpha_1 \cdot t) \cdot E_{ijt} + (\gamma_0 + \gamma_1 \cdot t) \cdot Q_{ijt} + Z_{ijt}\beta + \xi_i + \phi_j + \sigma_t + \omega_{ijt} \tag{4.1}$$

where I_{ijt}, or "internet", is an indicator of a consumer i's decision to shop online (rather than off-line) in chain j at time t; expenditure E_{ijt} is the observed expenditure in the trip; Q_{ijt} is the number of items, or "quantity", that she purchased; Z_{ijt} contains a vector of observables, including distance D_{ijt} to the closest outlet of chain j, and a vector of experience indicator $\mathbf{1}(X_{it} = x)$. The marginal effect vector β does not vary with time. We allow the marginal effect of expenditure and quantity to vary with time t, with a linear time trend, captured by α_1 and γ_1.

Alternatively, to test whether expenditure sensitivity varies with experience, we estimate a similar specification, where the response to expenditure and quantity contain a trend in experience but not in time:

$$I_{ijt} = c + (\alpha_0 + \alpha_1 \cdot X_{it}) \cdot E_{ijt} + (\gamma_0 + \gamma_1 \cdot X_{it}) \cdot Q_{ijt} + Z_{ijt}\beta + \xi_i + \phi_j + \sigma_t + \omega_{ijt}. \qquad (4.2)$$

The error term $\omega_{ijt} = I_{ijt} - \mathbb{E}[I_{ijt}|E_{ijt}, Q_{ijt}, Z_{ijt}]$. Because expenditure is a decision consumers make, consumer who face higher risk online will lower their expenditure. Therefore, there is a reverse causality issue when estimating specification 4.1. We estimate specifications (4.1) and (4.2) by instrumental variables regression with fixed effects. Specifically, we instrument E_{ijt} and Q_{ijt}, and their interactions with time or experience. The choice of instruments will be discussed in the Section 4.4.3.2.

In addition, when we take experience in the interaction term, we need to be mindful of the fact that experience is constructed as a summation over past decisions, or

$$X_{it} = \sum_{\tau<t} \sum_{j} I_{ijt}.$$

Therefore, the individual fixed effect ξ_i will enter X_{it}, and therefore Equation (4.2), nonlinearly. In addition, the number of periods or online trips one observes for an individual is also selected. To address this concern, we only use a sub-sample of individuals who have *no less than* 10 online trips in total, in their entire duration in the sample. And we focus on their online shopping choice when X_{it} is *no larger than* 10. Intuitively, we fix the set of individuals to be homogeneous across different values of X_{it}. In addition, instrumenting the interaction term $X_{it} \cdot E_{ijt}$ also alleviates some of the endogeneity concerns.[9]

[9]We never use instruments interacted with experience trend as instruments; instead, we use interaction between instrument and time trend as exogenous variations to endogenous variables interacted with experience.

Table 4.6: Variation of minimal distance

	households	share of hh	household-chain pairs	share hh-chain
(1) due to relocation	664	0.06	1,175	0.04
(2) due to entry/exit	3,122	0.29	5,032	0.15
either (1) or (2)	3,349	0.31	5,570	0.17
total observations	10,803		32,929	

Notes: This table reports percentages of households, or household-chain pairs with variations in the minimal distance.

4.4.3 Identification

4.4.3.1 Exogenous variations in travel distance

We identify the impact of potential travel cost on a consumer's decision between shopping online or in a store, from exogenous variation in the great-circle distance between a consumer's home address to the nearest outlet of a given chain. Conditional on time-invariant characteristics of consumers and chains (e.g. unobserved brand equity), the variation of distance comes from two sources: 1) relocation of a consumer, and 2) entry and exit of a store.

Table 4.6 reports share of the sample that has variation in household-chain minimal distance, and decomposes the source of variation into relocation and outlet entry and exit. On the household level, 3,349 households (31%) whose data contain variation in the distance to at least one chains; among which, 663 (6%) contain variation from household relocation, and 3,122 (29%) contain variation due to store entry and exits. These 31% households represent 17% observations in the household-chain pairs.

Table 4.7 checks representativeness of household-chain pairs, with and without minimum distance variation. We do not find economically-significant differences between the distance and expenditure per trip, tendency to shop online, and monthly net income. This suggests that the variation in distance in a sub-sample provides fairly representative estimates of the entire sample.

Table 4.7: Summary statistics of household-chains with and without distance variation

	with variation: mean	std err of mean	w/o variation: mean	std err of mean
distance (10km)	0.76	0.01	0.80	0.01
shop online (binary)	0.05	0.00	0.04	0.00
expenditure (10 euros)	3.52	0.03	3.74	0.02
netincome (1000 euros)	2.49	0.01	2.52	0.01
obs.	5570	5570	27993	27993

Notes: This table reports summary statistics of key variables by groups of household-chain pairs, with or without variations in the minimum travel distance.

4.4.3.2 Choice of instruments

Net income We postulate that the monthly net (post-tax) income that the household receives is exogenous to her online-shopping unobserved decision shocks; yet, households with higher net income generally spend more each trip. In the data, 35% households have income variation in the sample period. We also include the interaction term between net income and a time trend as instrument.

A potential concern that net income might be correlated with unobserved online shopping tendency, is that if net income variation is driven by unemployment, re-employment or retirement, the changes in the opportunity cost of time will alter an individual's preference for in-store shopping. We address this by controlling for employment and retirement status, and only use the variation in net income conditional on the employment status.

Share of products on deal We also instrument expenditure and quantity by the share of products on deal, for a given chain in a week. Since major chains in the Netherlands time their deals differently (despite common seasonality effects, which will be controlled for), variations in deal timings across chains can capture some variations in the consumer expenditure. Figure 4.3 shows an example of such, where the share of products on deal is different across major chains, despite some common seasonality effects.

On the other hand, the observed share of deal is based on the set of products purchased by the consumer, which will likely over-state the deal probability. This itself does not invalidate the use of deal as an instrument, as long as all correlations in the unobserved online shopping tendency and shocks of deal discovery (hence choosing products on deal) are captured by individual, chain and time effects. To provide a counter-example, if we do

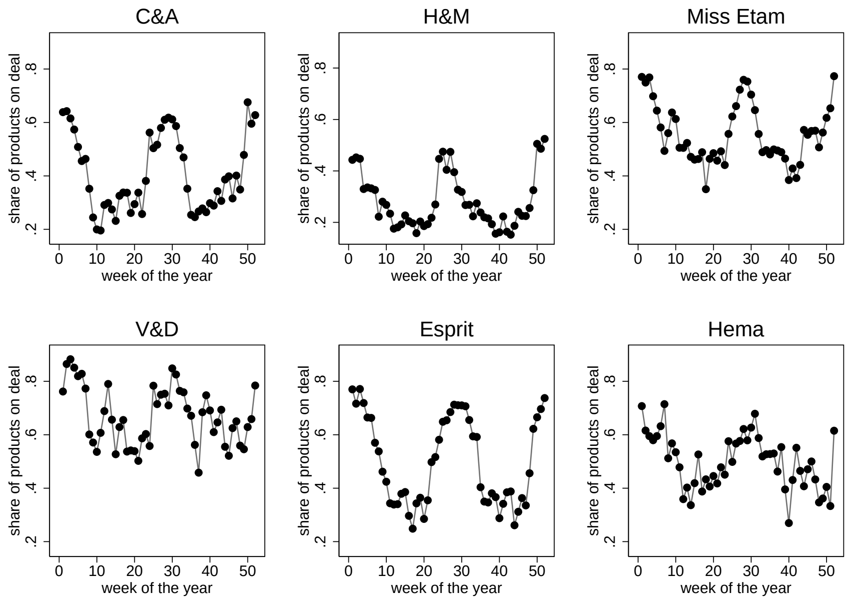


Figure 4.3: Timing of deals across chains

Note: This figure presents the average share of products on deal, by chain and week of the year. Cyclicality is present but different across chains.

not control for month fixed effects, then the mid-year spike in deal share might be correlated with seasonal patterns in online shopping. As a confirmation that we indeed control for all remaining correlations, Figure 4.4 documents the difference between the share of products on deal between online and in-store channel. We find that the distribution of difference in deal probability is very close to being symmetric around zero. This indicates that when individual, chain and month fixed effects are controlled for, online shopping decision shocks do not noticeably correlate with deal shares.

Product type In addition, we instrument expenditure and its time trend by the type of clothes that the individual purchased. We assume that the share of product type in a given trip (e.g. 2 T-shirts among the 3 items in a given trip) is driven by consumer need, and is unrelated to channel choice. If the shares of winter and summer clothes are excluded in the channel choice problem, they are good instruments since the expenditure varies across summer and winter products.

This assumption would be violated if unobserved weather drives consumer to buy winter

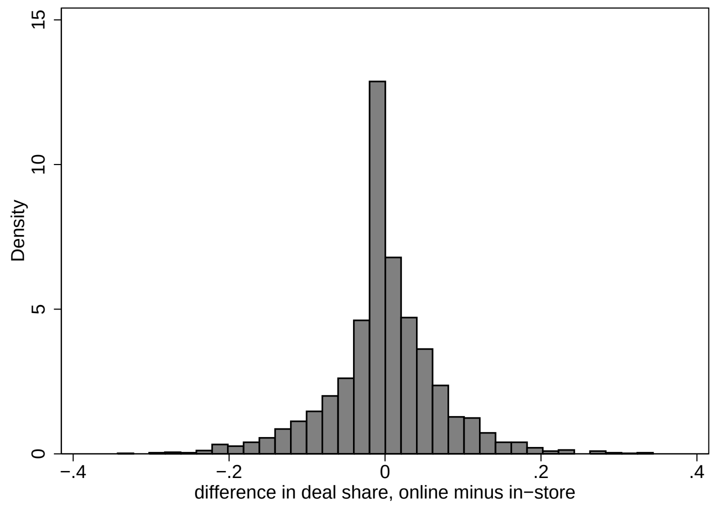


Figure 4.4: Difference in deal share between channels, within individual-chain-month

Note: This figure presents distribution in the difference in deal share across channels, at the individual-chain-month level data. Symmetry suggests that online channel does not systematically generate more sales with deals, when individual, chain and month fixed effects are controlled.

clothes, or if winter clothes are harder to carry – in both cases, purchase of winter clothes is associated with higher travel cost. To rule out this possibility, Figure 4.5 plots the choice of winter clothes and summer clothes by month of the year, along with usage of online shopping. Here, winter clothes are defined as product types of “coat”, “jacket”, “sweater”, “vest” or “spencer”, and summer clothes as “T-shirt”, “skirt” and “polo-shirt”. Individual and chain fixed effects, and weather fixed effects are controlled. We find that the cyclicality of internet is almost fully captured by weather and other control variables, whereas the variation in the choice of clothing displays very clear cyclicality.

4.4.4 Instrumental variable estimates

4.4.4.1 First stage

We estimate specifications (4.1) and (4.2) by two stage least squares (2SLS). That is, we separately estimate OLS of expenditure, quantity and their interaction term with time or experience, on excluded variables, covariates Z_{ijt} and individual, chain and time fixed effects. When estimating the specification (4.2) with experience, selection in the total number of

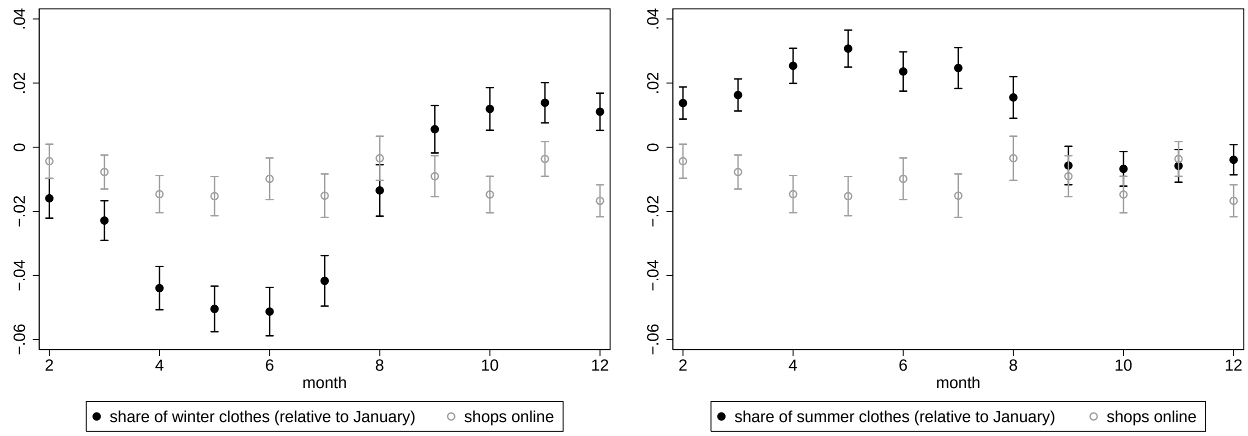


Figure 4.5: The use of internet and choice of winter clothes

Note: This figure presents OLS estimates of month fixed effects (relative to January), of the share of winter clothes ("coat", "jacket", "sweater", "vest" or "spencer") and summer clothes ("T-shirt", "skirt", "polo-shirt"), and the choice of shopping online. Weather, individual and chain fixed effects are controlled. These figures show that the cyclicality of internet is almost fully captured by weather, individual and chain fixed effects.

shopping trips requires that we focus on the first 10 trips for a sub-sample of consumers who shopped no less than 10 times online. Therefore, we also estimate first stage regressions using the sub-sample of 683 consumers. These are reported in the first two columns, while estimates from the full sample (10,794 consumers) are reported in columns 3-4.

In the first stage equation, we find that distance is positively correlated with expenditure and the number of units, which is consistent with a scale-economy argument, that a higher fixed cost for traveling further must be rationalized by a larger-scale purchases. The share of products on deal is negatively correlated with expenditure, but positively correlated with the number of units. This is consistent with a downward-sloping demand curve. The log monthly income parameter suggests that there are neither economically nor statistically insignificant effects from income on expenditure. The share of fall/winter clothes increases expenditure, while decreases the number of units. But we do not find significant effect on the share of spring/summer clothes.

Finally, we test for weak instruments, by jointly testing whether the coefficients of deal share, net income, the two product type variables, and their interaction with time trend, are jointly zero in the first stage. F-test statistics indicate that the instruments are weak in the sub-sample, but strong in the full sample. However, as shown by the second stage results,

Table 4.8: Regression of expenditure and quantity on instruments (first stage)

	expenditure: sub-sample	units: sub-sample	expenditure: full sample	units: full sample
distance (10km)	0.099*	0.019	0.105***	0.030***
	(0.058)	(0.033)	(0.019)	(0.010)
share of products on deal	-1.472***	0.492**	-1.285***	0.428***
	(0.375)	(0.216)	(0.133)	(0.070)
deal share* years	0.294***	0.027	0.130***	0.000
	(0.097)	(0.056)	(0.030)	(0.016)
log income	-0.156	-0.025	0.058	0.012
	(0.288)	(0.165)	(0.087)	(0.046)
log income* years	0.042	0.049	-0.025	0.000
	(0.065)	(0.037)	(0.016)	(0.009)
spring/summer clothes	-0.211	-0.062	-0.043	0.083**
	(0.199)	(0.114)	(0.068)	(0.036)
summer clothes* years	0.077	0.050	-0.052***	-0.009
	(0.064)	(0.037)	(0.018)	(0.010)
fall/winter clothes	0.417**	-0.182*	0.206***	-0.188***
	(0.191)	(0.110)	(0.066)	(0.035)
winter clothes* years	-0.017	0.009	0.009	0.008
	(0.054)	(0.031)	(0.016)	(0.008)
constant	4.147***	1.867***	3.585***	1.708***
	(0.612)	(0.352)	(0.214)	(0.113)
weather fixed effect	Yes	Yes	Yes	Yes
year-month fixed effect	Yes	Yes	Yes	Yes
chain fixed effects	Yes	Yes	Yes	Yes
nr trip fixed effect	Yes	Yes	Yes	Yes
employment/retirement status	Yes	Yes	Yes	Yes
Rsq.	0.182	0.036	0.165	0.018
obs.	13177	13177	125681	125681

Notes: This table reports first stage estimates of expenditure and quantity on instruments and exogenous covariates, as first stage estimates of the 2SLS on (4.1) and (4.2). Excluded instruments are: log of net monthly income, indicator of summer and winter clothes, share of products on deal, and their interaction with calendar time. The first two columns report estimates on a sub-sample, for individuals whose total online trip is no less than 10 times, and focus on their shopping decisions before the 11th online trip. For the sub-sample, test for the explanatory power of all excluded variables in the first stage yields F-statistics of 4.12 (expenditure) and 3.51 (quantity). For the full sample, test statistics are 27.04 (expenditure) and 23.49 (quantity). *, ** and *** represent significance at 0.10, 0.05 and 0.01.

in Table 4.9, the weak instrument still provide enough variation for statistical inference.

4.4.4.2 Second stage

Table 4.9 presents second stage estimates of Equations (4.1) and (4.2), using as instrumented variables: net income, deal share, and share of seasonal products, and their interaction terms with time. For consumers with no online-shopping experience,[10] we find that expenditure *causally* reduces online shopping tendency by 6.5%; in addition, each time the consumer shops online will increase the sensitivity to a unit change in expenditure in subsequent online shopping decision by 1.2%. The main effect on expenditure is insignificant at 95% confidence (two-sided p-value is 0.105), but the experience trend estimate is statistically significant, despite weak instruments.

On the other hand, for the full sample, we find that expenditure will lower online-shopping tendency, but the effect is insignificantly different from zero. The main reason is the smaller effect size, which might be due to that we are comparing across different sets of consumers. However, we do find that the marginal effect of expenditure is smaller (in absolute value) in later years: every year, the marginal effect of expenditure increases (decreases in absolute value) by 0.4%. We control for individual, chain, time, experience and employment status fixed effects.

Also, we find that each 10 kilometer in additional travel cost increases online shopping tendency by 1.4%, and the dis-utility in travel cost is very similar across consumers. This indicates that selection in past online shopping experience is *not* due to heterogeneity in travel cost.

4.4.4.3 Further robustness checks

In past versions, we estimated the two reduced-form specifications using log specifications on expenditure, quantity or distance. There were no drastic changes in parameter estimates that alters implication. Also, we allow the travel cost coefficient to depend on time, and do not find a statistically or economically meaningful time trend.

[10]But among the sub-sample of consumers who would ultimately have enough experience.

Table 4.9: Online shopping on distance and expenditure (second stage)

	online choice prob	online choice prob
expenditure (10 euros)	-0.065	-0.021
	(0.040)	(0.017)
expnd.*nr trips	0.012***	
	(0.005)	
number of items purchased	-0.027	0.004
	(0.060)	(0.024)
nr items* nr trips	-0.002	
	(0.010)	
distance (10km)	0.015*	0.014***
	(0.007)	(0.003)
expenditure*years		0.004***
		(0.001)
nr items* years		-0.006*
		(0.004)
constant	0.192	0.040
	(0.218)	(0.090)
weather fixed effect	Yes	Yes
year-month fixed effect	Yes	Yes
chain fixed effects	Yes	Yes
nr trip fixed effect	Yes	Yes
employment/retirement status	Yes	Yes
obs.	13177	125681

Notes: This table reports instrumental variable estimates of the choice of online shopping (I_{ijt}), on the total expenditure of the given trip, and the distance to the closest retail outlet. Expenditure, quantity and their interaction with time are instrumented by the log net monthly income, share of products on deal, and share of purchases with summer or winter clothes, and their interaction with time trend. The first column reports estimates on a sub-sample, for individuals whose total online trip is no less than 10 times, and focus on their shopping decisions before the 11th online trip. *, ** and *** represent significance at 0.10, 0.05 and 0.01.

4.4.5 Discussion: Does online-purchase risk explain the increasing channel substitution?

Prince (2007) documents increasing substitution among traditional and online retailers, and attribute this to consumer and (mainly) retailer heterogeneity. Melis et al. (2014) document that past online shopping experience (in their case, experience is defined as sum of expenditure online in a moving window) increases substitution between different retailers. Our reduced form analysis adds to the literature in two ways:

First, we emphasize within-retailer substitution between channels. While both Prince (2007) and Melis et al. (2014) emphasize changes in substitution over time or experience, they give emphasis on across-retailer substitution. In fact, Prince (2007) stresses substitution between traditional retailers and new retailers who only operates online; and Melis et al. (2014) emphasize substitution across web-shops of multi-channel retailers. However, to answer a managerial question on whether a firm should open its online branch, or a policy question on whether the online branches of existing retailers should be encouraged or regulated, we need insights that is specific to within-retailer substitution patterns.

Second, and most importantly, we propose a specific, within-consumer mechanism that drives the increasing channel substitutability. In fact, Prince (2007) proposes that the evolution of consumer and retailer *distribution* is the explanation of the substitution between retailers, and emphasizes store entry of the rapidly growing online business. On the other hand, Melis et al. (2014) document that a consumer's past cumulative online expenditure is positively related to the importance of product assortment in their current choice, but do not explicitly point to causality. We propose that changes in the perceived risk is one important mechanism (controlling for other mechanisms) that drives the increasing substitution patterns.

Finally, we document that changes in channel substitution occur both along calendar time, and along individual consumer's past online shopping experience. Even though we have not yet formally tested between the two,[11] our results hint that the proposed model

[11] Specifically, within the sub-sample, the instruments we use do not provide enough variation to test between time trend and experience.

should distinguish between time effects, which could occur if there are supply side changes in online service quality, or word-of-mouth type of information spillover *across* consumers; and experience effects, which could occur if there is learning.

4.5 A structural model of increasing channel substitutability

4.5.1 Overview

Following the discussion in Section 4.4.5, we propose a model that characterizes a consumer's choice between shopping online and off-line. The model incorporates learning and possible time effects, which generate increasing substitutability in the choice of two channels, over experience and time. The model is also very simple to solve.

We then fit the model onto a set of choice "patterns" generated from the data. A "pattern" is an estimate of conditional choice probability of online shopping – net of individual, chain and store fixed effects – as a function of an individual's experience, calendar time, and the distance to the closest outlet. We compute average conditional online choice probability, and similarly, conditional average expenditure, using data among households with no less than 10 total online trips.[12] We fit the model on choice probabilities and expenditure *net of* fixed effects, to obtain estimates free of selection on unobserved heterogeneity. This reduced-form control of heterogeneity helps maintain model simplicity.

4.5.2 Model setup

Our model characterizes a household i's choice between two channels, in chain j at trip t. For notational simplicity, we suppress i and j unless there is a special emphasis. Before the trip, the household thinks about a *shopping need*, and chooses between shopping online or "off-line", in the latter case she travels to the nearest outlet. Denote the choice $I_t = 1$ for online, and $I_t = 0$ for off-line trip.

Regardless of the channel choice, the household believes that there is a probability that

[12] This is the same sub-sample used in Column 1, Table 4.9.

the product turns out to be "bad". For example, a product is "bad" if its quality (or the service quality attached to it) turns out to be less than the advertised quality. In this case, the product yields utility zero.

If the consumer makes the purchases off-line, she observes the product quality before having to pay for it. If she purchases online, however, her expenditure is sunk before product quality is realized.[13] This is to say, the *expected* utility (i.e. utility before the trip) is discounted by the perceived probability of not receiving a bad product:

$$u_{1t}(E_{1t}, S_t) = (1-\delta_t)\cdot v(E_{1t}, \mu_t) - E_{1t}, \tag{4.3}$$

where $v(E_{1t}, \mu_t)$ is the *ex-post* consumption utility, realized when the product is *not* bad. In particular, μ_t denotes consumption utility shocks, and the consumer observes the realization of μ_t before her choice of channel. One can think about μ_t as the "shopping need", thought about before making the trip decision. $\delta_t \in (0,1)$ is the *perceived* probability that the product is bad; and the last term is dis-utility from expenditure, price coefficient normalized to 1. Finally, to simplify notation, we denote state vector $S_t = (\mu_t, \delta_t, D_t)$, which summarizes all relevant state variables: consumption preference shock, perceived risk and distance.[14]

On the other hand, if shopping off-line, her expenditure is decided after observing the true quality. To impose that she has consistent belief on the quality distribution, the consumer perceives that the probability of finding a bad product in store is the same as doing so online. Therefore, although her expenditure is not sunk before knowing the quality, her travel cost is sunk, and her expected utility from purchasing off-line is

$$u_{0t}(E_{0t}, S_t) - \varepsilon_t = (1-\delta_t)(v(E_{0t}, \mu_t) - E_{0t}) - f(D_t) - \varepsilon_t, \tag{4.4}$$

where the dis-utility from expenditure does not incur if the product is bad (with probability δ_t), and $f(D_t) + \varepsilon_t$ is a stochastic travel cost function on distance D_t, and random (logit) travel utility ε_t. Note that the term u_{0t} only denotes off-line utility net of $-\varepsilon_t$.

[13] For simplicity, we do not allow for the possibility of return.

[14] Note that online utility u_{1t} is degenerate on D_t.

4.5.3 Optimal expenditure

Despite having the same consumption utility function, $v(., \mu_t)$, optimal expenditure online and off-line are different, because of the additional uncertainty from online purchase. We give the consumption utility a quadratic functional form,

$$v(E_{\iota t}, \mu_t) = (\beta + \mu_t) E_{\iota t} - \frac{\gamma}{2} E_{\iota t}^2,$$

for $\iota = 0, 1$, in order to obtain closed-form solution to the optimal expenditure. We impose that β, $\gamma > 0$, so that consumption utility is increasing when expenditure is reasonably low, and decreasing after satiation is reached. In addition, the consumption utility function is concave, and the first order condition finds maximum utility. Substitute this specification into Equations (4.3) and (4.4) and take the first order condition with respect to expenditure, we have the optimal expenditure in an online shopping trip:

$$E_{1t}^* = \begin{cases} \gamma^{-1} \left(\beta - (1 - \delta_t)^{-1} + \mu_t \right) & \text{if } \mu_t > - \left(\beta - (1 - \delta_t)^{-1} \right) \\ 0 & \text{otherwise;} \end{cases}$$

and the off-line counterpart:

$$E_{0t}^* = \begin{cases} \gamma^{-1} (\beta - 1 + \mu_t) & \text{if } \mu_t > -(\beta - 1) \\ 0 & \text{otherwise.} \end{cases}$$

There are two remarks. First, optimal consumption will never go beyond the satiation point, because if that is the case, the consumer can always spend less while attaining at least the same utility level.

Second, because of the non-zero probability of getting a bad product, online expenditure is smaller than off-line, all else equal.[15] This prediction is confirmed in the empirical findings of Ansari et al. (2008). Intuitively, this is because an item is *effectively more ex-*

[15] To see this, compare terms and note that $(1 - \delta_t)^{-1} > 1$.

pensive if purchased online, because it requires $(1-\delta_t)^{-1}$ times the expenditure one would otherwise spend off-line, in order to receive a “good” product. Therefore, in optimality, a consumer will be more conservative when shopping online.

4.5.4 Choice probability and off-line expenditure

Substitute the optimal expenditure schedule conditional on channel, into the utility function, and one obtains the indirect utility from shopping online and offline, respectively, $u_{1t}(E_{1t}^*, S_t)$ and $u_{0t}(E_{0t}^*, S_t)$. Because of the logit assumption on ε_t, conditional choice probability follow standard logistic choice probability expression:

$$\Pr(I_t = 1|S_t) = \frac{\exp(u_{1t}(E_{1t}^*, S_t))}{\exp\left(u_{1t}\left(E_{1t}^*, S_t\right)\right) + \exp\left(u_{0t}\left(E_{1t}^*, S_t\right)\right)}.$$

However, the consumption utility shock μ_t is unobserved; hence, we are more interested in characterizing the average choice probability over the distribution of μ_t. We should also condition on positive expenditure in at least one channel, because only when the household has something to buy, the channel choice problem can correspond to actual purchase decisions. Therefore, the channel choice probability conditional on positive expenditure is

$$\hat{P}(\delta_t, D_t) \equiv \int \Pr(I_t = 1|S_t)\, dG(\mu_t|\mu_t > -(\beta - 1)), \tag{4.5}$$

where $G(.)$ denotes a general (joint) distribution function. Note that we used the property that the online expenditure is always lower than the off-line counterpart; so requiring optimal spending to be positive in at least one channel is to require positive optimal offline spending.

Finally, we match the model-computed expenditure in the off-line channel, to its data counter-part.[16] The model predicted off-line expenditure is the average of optimal expenditure, under two conditions. First, the underlying optimal expenditure should be positive ($\mu_t > -(\beta - 1)$). Second, the consumer chooses to shop off-line given the taste shock

[16] We do not match online expenditure because data on online purchase is relatively rare.

$(u_{0t} - \varepsilon_t > u_{1t})$.

$$\hat{E}(\delta_t, D_t) = \int E^*_{0t} dG(\mu_t, \varepsilon_t | \mu_t > -(\beta - 1), u_{0t} - \varepsilon_t > u_{1t}). \tag{4.6}$$

4.5.5 Parametrization

4.5.5.1 Perceived risk

We parametrize δ_{it} (here we emphasize that the perceived risk is individual and time specific) as a function of time t and experience X_{it}:

$$\begin{aligned} \delta_{it} &= \delta(t, X_{it}) \\ &= \sum_x \delta^x \cdot \mathbf{1}(X_{it} = x) + \sum_\tau \delta^\tau \mathbf{1}(t = \tau). \end{aligned}$$

Note that this specification is flexible, in the sense that there is no parametric assumption that restricts the form of learning or any other advancement. In model implementation, X_{it} takes integer values from 0 to 10, and time is discretized to two-year intervals: t takes values 2008, 2010, 2012 and 2014. We normalize δ^{2014} – the parameter for the last year grid – to 0.

4.5.5.2 Travel cost

We impose a quadratic specification on the cost function:

$$f(D_t) = f_0 + f_1 D_t + f_2 D_t^2.$$

The intercept term f_0 capture the innate utility to shop online (relative to off-line), irrespective of distance and expenditure. In other words, f_0 is the choice intercept.

4.5.5.3 Distribution assumptions

We normalize the location of expenditure shock μ_t, but do not normalize the scale. We assume that $\mu_t \sim \mathcal{N}(0,\sigma)$. The scale of ε_t is normalized due to the logistic distribution assumption.

4.5.6 Data moments and estimation sub-sample

Our aim is to match the predicted moments – choice probability $\hat{P}(\delta_t, D_t)$ as defined in (4.5), and offline expenditure $\hat{E}(\delta_t, D_t)$ as defined in (4.6) – onto the data. Note that in the previous section, we defined the perceived risk δ_t as a function of experience and time. Therefore, to estimate model parameters, we can directly match model prediction as a function of time, experience and distance, to data averages of choice probability and off-line expenditure. Specifically, *fixing* $X_{ijt} = x$ and t, we estimate

$$I_{ijt} = \bar{I}(t, x, D_{ijt}) + \alpha_{ij} + \nu_{ijt} \tag{4.7}$$

and obtain the average online shopping choice probability profile $\bar{I}(t,x,d)$ over different values of distance $D_{ijt} = d$. We control for individual and store combined fixed effects, which characterize individual and chain unobserved characteristics, as well as potential heterogeneity in the individuals' innate match values to different chains. Similarly, we can estimate the average expenditure profile $\bar{E}(t,x,d)$.

Note that we do not estimate $\bar{I}(t, X_{ijt}, D_{ijt})$ as a function of time and experience X_{ijt}. This is because X_{ijt} is constructed as past realization of online shopping decisions, and thus contains summation of individual fixed effects.[17] In addition, because X_{ijt} enters nonlinearly, the presence of fixed effects α_{ij} cannot be differenced out in a standard fixed effect linear model. This requires that we exploit variations in X_{ijt}, by only comparing across individuals with similar fixed effects. We hence focus on the sub-sample where each consumer shops online for at least 10 times, and only focus on their first 10 times of online shopping

[17] Because $X_{ijt} = \sum_{\tau<t}\sum_j I_{ij\tau}$, hence experience contains summation of individual-store fixed effects and error term over periods $\tau = 1, ..., t-1$ and over all chains $j = 1, ..., J$.

trips. Therefore, we always compare the average channel choice probabilities among the same set of consumers. On top of sample selection, controlling for linear fixed effects help correct for additional heterogeneity across individuals and chains.

This sample select criterion focuses on 875 households out of 14,989 households in total. However, many households in the full sample are "inactive" in that they contribute very few purchase records. We find that 4,456 households stayed in the data for at least 3 years, and each contain at least 20 purchase records (on the trip-retailer level). Therefore, our sub-sample is about a quarter of the more meaningful sample of households. On the other hand, these are households that like online purchase more than the rest. Our sub-sample of households contain 34,460 trip records, which is 17% of the entire data-set.

4.5.7 Estimation algorithm

For each candidate parameter, the model solves for the optimal expenditure online and off-line, and then gives predicted online choice probability $\hat{\Pr}(I_t = 1|t,x,d)$ and predicted average expenditure off-line, $\hat{E}(t,x,d)$. The model fit is then defined by a squared-distance criterion:

$$\mathscr{D} = \sum_{t,x,d} \left(w_{t,x,d} \left(\hat{\Pr}(I_t = 1|t,x,d) - \bar{I}(t,x,d) \right)^2 + w^0_{t,x,d} \left(\hat{E}(t,x,d) - \bar{E}(t,x,d) \right)^2 \right)$$

where $w_{t,x,d}$ and $w^0_{t,x,d}$ are sample weights, defined as the number of observations in each grid point when estimating (4.7) and the off-line expenditure counterpart. We then find parameter θ that minimizes the sum of squared distance $\mathscr{D}$.

To compute the model predicted choice probability and expenditure, we simulate decisions of 1,000 individuals, over 11 periods. For each individual, we exogenously give her a distance to a "typical" chain, and hold distance as fixed. For each individual in each period, we randomly generate a year indicator, drawn uniformly in $\{2008, 2010, 2012, 2014\}$. Since the choice problem is purely static, we do not require the "year" variable to be sequential. Finally, for each individual in each period, we take 100 draws of consumption utility shock μ_t and travel cost shock ε_t. We simulate choices and expenditure based on individual, time, and draw. Finally, we collect all observations with the same distance, year and experience grid, and compute the average model-predicted choice probability and expenditure. The utility shock and travel shock draws are fixed for all function evaluations of $\mathscr{D}$, and also for all bootstrap sample estimation – explained below.

We then take 50 bootstrap samples, each randomly drawn from the original data with replacement. For each bootstrap sample $b = 1,...,50$, we compute choice and expenditure profiles $\bar{I}^b(t,x,d)$ and $\bar{E}^b(t,x,d)$. Then, we find parameter that minimizes the sum of squared distance of model prediction and data profiles from the bootstrap sample, which gives us a separate set of estimates θ^b for each b. Finally, standard error of the estimates are computed as standard deviation over the entire set of estimates $\left\{\theta^b | b = 1,...,50\right\}$.

4.6 Results

4.6.1 Parameter estimates

Table 4.10 presents parameter estimates and the bootstrap standard errors. For the components of the perceived risk $\delta(t,x)$, we find that experience has a strong and significant effect of reducing (in general) the perceived risk, compared to the effect of time. Consumers who has no prior online shopping experience believes that online shopping is 20% riskier, compared to the cases when they have had 1-3 online trips before. Beyond that, additional experience further drives down perceived risk: consumers with 4-10 past online trips have perceived risk that is 10-15% lower than those with 1-3 trips in the past. We do not find a strictly monotonic relationship between past online shopping experience and current perceived risk. However, it is difficult to conclude whether the (slightly) noisy estimates at the tail is due to sampling error, or heterogeneity in the past shopping experiences among different consumers who has the same number of past online trips. A potential model for the latter explanation would be a Bayesian learning model for mean belief: posterior mean is the weighted average of past signals, hence, variance of the posterior mean across consumers increase in the length of history.

We graphically present the experience effect in the estimated δ_t, in Figure 4.6. We

Table 4.10: Estimates of structural parameters

	parameter	std err
online discount parameter: no online trip before	0.41	0.08
– 1 trip before	0.20	0.09
– 2 trips before	0.20	0.08
– 3 trips before	0.18	0.07
– 4 trips before	0.08	0.10
– 5 trips before	0.03	0.08
– 6 trips before	0.03	0.09
– 7 trips before	-0.01	0.07
– 8 trips before	0.02	0.08
– 9 trips before	0.04	0.06
– 10 trips before	0.09	0.08
– in years 2007-2008	0.12	0.16
– in years 2009-2010	0.19	0.18
– in years 2011-2012	0.06	0.37
consumption function: expenditure coef. (β)	16.41	1.84
– expenditure squared (γ)	4.33	0.52
fixed cost function: intercept (f_0)	-1.92	0.14
– distance (f_1)	1.35	0.37
– distance squared (f_2)	-0.30	0.44
standard deviation for consumption utility shock (σ)	6.82	0.54

Notes: This table reports parameter estimates of the structural model, explained in Section 4.5. Standard error computed from 50 bootstrap reps. As a scale normalization, note that distance is defined in units of 10km, and expenditure defined in units of 10 euro.

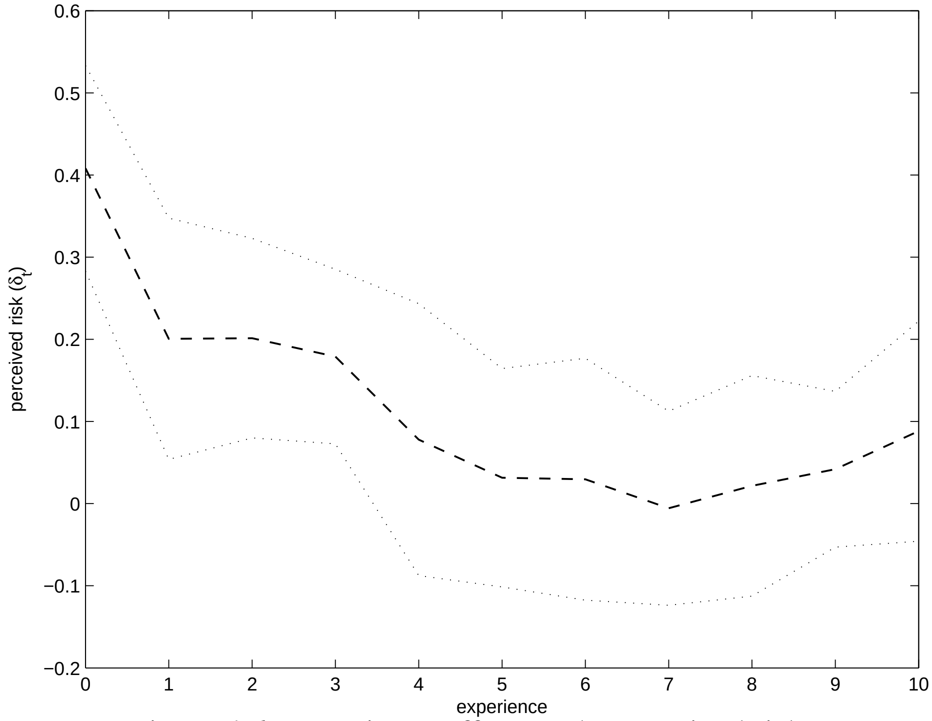


Figure 4.6: Experience effect on the perceived risk

Note: This figure presents estimates of the effect of experience on the perceived risk: δ^x for $x = 0, ..., 10$. Confidence intervals are computed using the bootstrap standard errors.

also plot the confidence intervals, which are computed by the bootstrap standard error of corresponding parameters. Initial experience has the largest effect, and the entire experience curve is very close to a concave curve, which can be rationalized by standard Bayesian learning theory (in the mean).[18]

As mentioned before, experience effects dominate time effects. We find that in years 2011-2012, perceived risk is about 5-10% lower than before 2010. However, this comparison is statistically insignificant from zero. Therefore, we find suggestive but not solid evidence that trips in later years reflect belief of lower online-shopping risk.

Simultaneously, we estimate a consumption utility function on expenditure. We find that consumption utility is concave. For the average consumer, consumption utility reaches its maximum at a moderate expenditure level (38 euro, slightly higher than the observed aver-

[18] A Bayesian learning model in the variance will fail to account for the increase in variance of δ_{it}.

age expenditure).[19] Also, a large part of the variations in observed expenditure is explained by variation in the utility shock, which has a standard deviation of 6.82.

The consumer also incurs travel cost when shopping off-line. We estimate a quadratic travel cost specification, where the intercept, f_0, is estimated to be negative. This reflects some innate travel preference towards shopping off-line, which might correspond to utility from other activities associated with shopping, such as going outside, family event, and so on. In addition, we find that traveling in longer distances incur increasing dis-utility,[20] but at a decreasing rate. This indicates that additional travel distance seem less costly, which might reflect endogenous changes of travel means. For example, a consumer might decide to drive if she goes beyond 10km, which might lower the additional travel dis-utility. Finally, note that distance is not the only travel cost; ε_{it} also causes variations in actual travel decisions.

4.6.2 Model fit

Figure 4.7 presents model fit. For each of the top left, top right and bottom left panels, we present model fit of average choice probability profile, in contrast to the data-generated choice probability *profile*, separately by distance, experience and time. For example, to produce the top left panel, we aggregate all model-simulated choices that fall into each grid point of distance, weighted by their frequency in the simulated data; and this produces the model-predicted choice probability. On the other hand, the data average choice profile is computed as frequency-weighted average of the conditional choice probability $\bar{I}_{t,x,d}$, generated by the estimation procedure in Section 4.5.6. The top right and bottom left panels are produced using similar methods, but conditional on experience or year.

For the three model fit plots of average choice probability, we find that the simple model does a good job in fitting the data. The quadratic functional form in distance keeps good track of the average distance response profile in the data, and the flexible specification in δ_t allow the model to keep track of evolution of choice probabilities in experience and time.

[19] Note that *mean* consumption utility is defined as

$$v = \beta E + \frac{\gamma}{2} E^2.$$

Also note that we define expenditure in units of 10 euro, as we do in the reduced form analysis.

[20] Until a distance of 22.5 km, which is almost at the edge of the observed distance domain.

In particular, the shape of choice probability in experience closely resembles the shape in Figure 4.6.

On the other hand, the bottom right panel shows model fit of the simulated off-line expenditure *distribution*, to the observed expenditure distribution without correcting for any fixed effects. Other than that the model-generated mean expenditure is still close to the data average, the expenditure distribution is fitted poorly. There are two reasons for that. First, we have not corrected for any fixed effects and errors, and doing so will take away a large part of the variance in expenditure raw data. As a result, the model-predicted expenditure contains lower variance than the data counterpart. Secondly, we have not imposed any higher-order moment condition in estimation (and there is no clear second moment condition to impose), so the model only resembles data mean but not data variance. This last task informs us of one limitation of this estimation approach: that we should not use this result to predict the *non-systematic* variability of expenditure.

4.6.3 What happens if consumers had more experience?

We simulate the counterfactual scenario, assuming that consumers had maximum possible experience, so that the online discount parameter only depends on calendar time effect: $\tilde{\delta}_{it} = \sum_{\tau} \delta^{\tau}$. That is, we impose all experience components in the perceived risk to be zero. We then simulate counterfactual choice probabilities, replacing model estimate δ_{it} with $\tilde{\delta}_{it}$.

Figure 4.8 presents the counterfactual outcome against experience and distance. It is intuitive that consumers go online more often, and against experience, the optimal choice probability reaches its maximum around 0.25. That is, if an inexperienced consumer is to be fully informed about online shopping, her online shopping tendency would have been *tripled*. We also plot the choice probabilities against distance, for consumers with zero experience and for those who are fully informed. Online shopping tendency increase under all distance, but the effect is much larger with greater distance. This implies that online experience increases channel substitutability sharply.

Because consumers with more experience spend more trips online, it saves them travel cost. We also compute the average reduction in consumer travel distance, from Figure 4.8 right panel and the empirical distribution of distance. We find that on average, a consumer

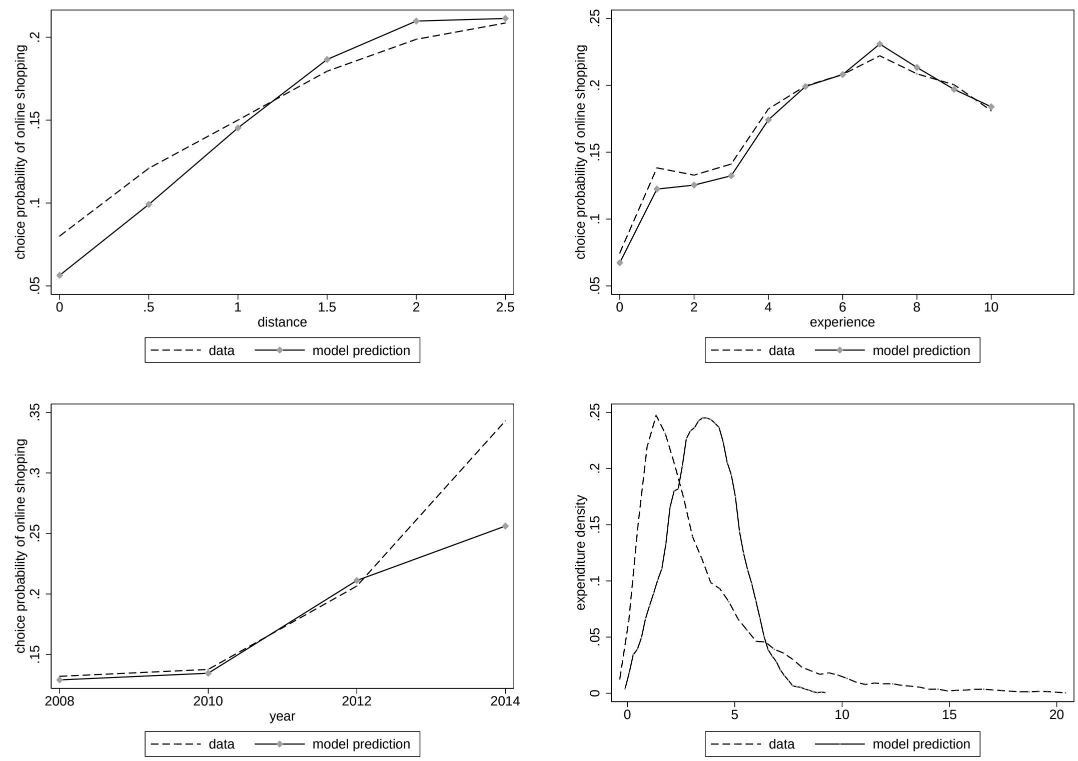


Figure 4.7: Model fit

Note: These four figures present model fit. The top left panel is the average online shopping choice probability, conditional on different grid points of distance (in 10 km). To plot this figure, we aggregate all model-simulated choices that fall into each grid point of distance, weighted by the frequency in the simulated data – this is the "model prediction". Data average is computed as frequency-weighted average of conditional choice probability $\bar{I}_{t,x,d}$, in each grid point of d. The top right and bottom left panel present fit of choice probability conditional on experience and year, respectively. They are computed using similar method. Finally, the bottom right panel is observed expenditure *distribution* – not corrected for any fixed effects – and model-predicted expenditure *distribution*.

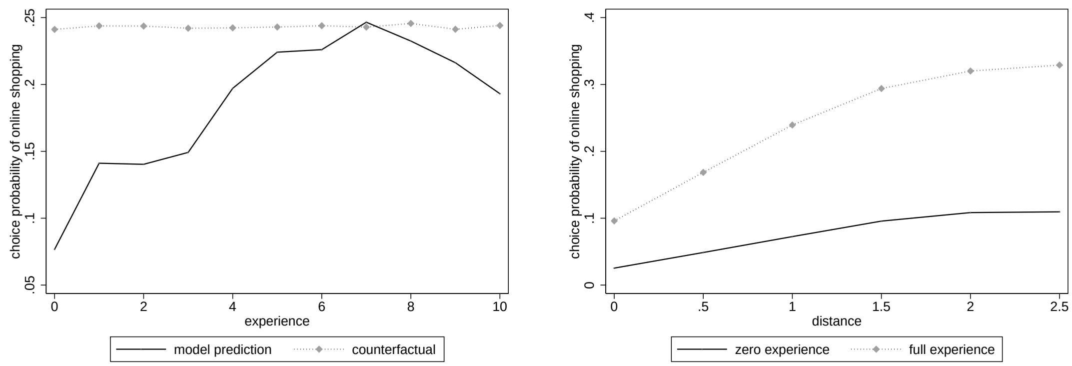


Figure 4.8: Counterfactual channel choice probability when consumers are fully experienced

Note: The left panel presents counterfactual channel choice probability against experience and distance, under the counterfactual scenario that all consumers have the highest potential experience regarding online shopping. Technically, we set $\delta^x = 0$ for all x. The right panel presents choice probabilities against distance, for consumers with zero experience and their δ^0 as estimated (solid line), versus when we set their $\delta^0 = 0$.

travels 1.27 km less for *each trip*.

4.6.4 Managerial implications

4.6.4.1 Quantifying cannibalization effects

We find that the online and off-line retail channels are substitutable for a given consumer, and their substitutability increases over time and experience. This implies that cannibalization between the two channels, for a given retailer, i) exists and ii) is increasing over time.

In the literature, channel choice sensitivity is usually confounded with store choice sensitivity (Prince 2007; Forman et al. 2006; among others). The amount of heterogeneity across retailers will result in small across-retailer sensitivity, but this gives the wrong impression that demand is also insensitive to distance *within* a retailer. In addition, in practice, within-consumer cannibalization effects are often under-emphasized, if not completely ignored. To give one example, a practitioner discussion about whether a major food retailer should go online or not, claims that (as cited by Melis et al., 2014): "*If [Morrisons] don't [enter the online food shopping business], Tesco, Sainsbury's and Asda will be taking shoppers*

[Morrisons] could have earned".[21] Claims like this emphasize only on business stealing effects across retailers in the same channel, implicitly treating online and offline consumers as completely isolated groups.

In this paper, we find large cannibalization effect within a retailer. In fact, for an experienced consumer living at the average distance, opening a new outlet at her doorstep would take away 10% sales from the online store. Conversely, if there were no online stores, opening one will cannibalize part of the off-line sales, apart from any potential brand-expansion effects.[22]

In addition, since channel substitutability increases with consumer experience and calendar time, so is the cannibalization effect. This further puts caution to retailers who are *currently* considering opening an online store, because (as we show) a large part of the high online shopping demand comes from the offline counterpart.

4.6.4.2 Experience and online shopping decisions

In this paper, we offer a new explanation to why online and off-line channels are increasing substitutes. Specifically, we claim that increasing substitutability between retail channels is caused by changes in online-shopping risk over time and experience, which is shown by reduced form evidence and structural estimates. In particular, we find that experience effect in reducing the perceived risk is much larger than calendar time effects.

Calendar time effects could be better websites or service quality, or more favorable product return guarantees, or improvement in regulation from the authorities; or, it could represent word-of-mouth effects on the demand side. On the contrary, a consumer's own experience effect can only be *subjective*, so as to rationalize why there are large and systematic differences between the online shopping tendencies of two consumers at the same time. The subjective difference then implies that inexperienced consumer over-estimates her potential risk of not getting what she intended to.

Although this can be fixed by having more online experience, it also implies potential for policy intervention. For example, a retailer (whose objective is to increase online sales)

[21] Melis et al. (2014) page 2.

[22] In our empirical exercise, we do not quantify the expansion of the total sales of a given chain.

can price penetrate the demand side, by offering exclusive discounts on the online shop. A consumer driven by such discounts will be gradually more open to shopping online, and the overall reduction in her shopping cost (i.e. reduced uncertainty cost, translated into reduced travel cost) implies that she will purchase more in the future.

4.7 Concluding remarks

When shopping cost varies, the way that a consumer substitutes between shopping online and off-line, changes over time. This paper aims at understanding *how* and *why* this is the case. With respect to "how", we provide descriptive evidence that the sensitivity to distance in a consumer's channel choice is 3 times as high in 2014, than in 2007; also, it is 3 times as high when a consumer has shopped 10 times online in the past, than when she had no or little experience online. This indicates that online and off-line channels are increasing substitutes in calendar time and/or consumer experience.

With respect to "why", we propose that a consumer's perceived probability of receiving an inferior product (hence the term "perceived risk") decreases over time. Because a consumer shopping in the traditional outlet can verify product quality before the purchase, off-line shopping naturally serves as an "insurance" against the risk of inferior product, and the travel cost associated with off-line shopping can be viewed as an insurance premium. When the perceived risk goes down, consumers are less reliant on verifying product quality before hand; hence, their choices are more sensitive to variations in travel cost. We test whether an exogenous shift in expenditure decreases online-shopping tendency, and find strong support for the hypothesis.

We then structurally characterize a consumer's choice whether to shop online and her expenditure given the choice, by constructing a simple model while controlling for a rich set of individual, chain and time fixed effects. Our model numerically solves for average online-shopping tendency and expenditure, as functions of calendar time, experience and distance. Meanwhile, we non-parametrically characterize the average choice and expenditure profiles (controlling for fixed effects), and match these to the model predictions. We find that past online-shopping experience contributes the stronger effect, than calendar time.

With the first three times shopping experience online, a consumer would find online shopping safer, in terms of a 20%-lower perceived risk, than what she believed when she had no experience. Further experience decreases the perceived risk by another 10-15%. Comparatively, evolution of the market environment (for instance, better websites or return services) during 2007-2014 decreases the perceived risk by 12%. This indicates that changes within a consumer is the main explanation of why the online stores of given retailers slowly gain more market shares.

Bibliography

Adner, Ron and Daniel Levinthal (2001), 'Demand heterogeneity and technology evolution: implications for product and process innovation', *Management science* **47**(5), 611–628.

Alba, Joseph W and J Wesley Hutchinson (1987), 'Dimensions of consumer expertise', *Journal of consumer research* pp. 411–454.

Albuquerque, P. and Y. Nevskaya (2012), 'The impact of innovation on product usage: A dynamic model with progression in content consumption'.

Albuquerque, Paulo and Bart J. Bronnenberg (2012), 'Measuring the impact of negative demand shocks on car dealer networks', *Marketing Science* **31**(1), 4f–23.

Allenby, Greg M, Thomas S Shively, Sha Yang and Mark J Garratt (2004), 'A choice model for packaged goods: Dealing with discrete quantities and quantity discounts', *Marketing Science* **23**(1), 95–108.

Ansari, Asim, Carl F. Mela and Scott A. Neslin (2008), 'Customer channel migration', *Jounal of Marketing Research* **45**(1), 60–76.

Arellano, Manuel and Stephen Bond (1991), 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The review of economic studies* **58**(2), 277–297.

Becker, Gary S and Kevin M Murphy (1993), 'A simple theory of advertising as a good or bad', *The Quarterly Journal of Economics* pp. 941–964.

Becker, G.S. (1965), 'A theory of the allocation of time', *The economic journal* **75**(299), 493–517.

Bell, David R, Jeongwen Chiang and Venkata Padmanabhan (1999), 'The decomposition of promotional response: An empirical generalization', *Marketing Science* **18**(4), 504–526.

Bell, David R. and Sangyoung Song (2007), 'Neighborhood effects and trial on the internet: Evidence from online grocery retailing', *Quantitative Marketing and Economics* **5**(4), 361 – 400.

Benkard, C Lanier (2000), 'Learning and forgetting: The dynamics of aircraft production', *American Economic Review* pp. 1034–1054.

Besanko, David, Ulrich Doraszelski, Yaroslav Kryukov and Mark Satterthwaite (2010), 'Learning-by-doing, organizational forgetting, and industry dynamics', *Econometrica* **78**(2), 453–508.

Bettman, James R, Mary Frances Luce and John W Payne (1998), 'Constructive consumer choice processes', *Journal of consumer research* **25**(3), 187–217.

Bronnenberg, Bart J, Jean-Pierre H Dube and Matthew Gentzkow (2012), 'The evolution of brand preferences: Evidence from consumer migration', *American Economic Review* **102**(6), 2472–2508.

Bronnenberg, Bart J, Michael W Kruger and Carl F Mela (2008), 'Database paper-the iri marketing data set', *Marketing Science* **27**(4), 745–748.

Bucklin, Randolph E., Jorge Silva-Risso and S. Siddarth (2008), 'Distribution intensity and new car choice', *Journal of Marketing Research* **45**(3), 473–487.

Ching, Andrew T, Tülin Erdem and Michael P Keane (2013), 'Invited paper-learning models: An assessment of progress, challenges, and new developments', *Marketing Science* **32**(6), 913–938.

Chintagunta, Pradeep K (1993), 'Investigating purchase incidence, brand choice and purchase quantity decisions of households', *Marketing Science* **12**(2), 184–208.

Chintagunta, Pradeep K., Junhong Chu and Javier Cebollada (2012), 'Quantifying transaction costs in online and offline grocery channel choice', *Marketing Science* **31**(1), 96–114.

Choi, Jeonghye, Sam K. Hui and David R. Bell (2010), 'Spatio-temporal analysis of imitation behavior across new buyers at an online grocery retailer', *Journal of Marketing Research* **47**(1), 65 – 79.

Clarkson, Joshua J, Chris Janiszewski and Melissa D Cinelli (2013), 'The desire for consumption knowledge', *Journal of Consumer Research* **39**(6), 1313–1329.

Crawford, G.S. and M. Shum (2005), 'Uncertainty and learning in pharmaceutical demand', *Econometrica* **73**(4), 1137–1173.

Dehmamy, Keyvan and Thomas Otter (2014), 'Utility and attention-a structural model of consideration', *Available at SSRN 2433145* .

Deleersnyder, Barbara, Inge Geyskens, Katrijn Gielens and Marnik G Dekimpe (2002), 'How cannibalistic is the internet channel? a study of the newspaper industry in the united kingdom and the netherlands', *International Journal of Research in Marketing* **19**(4), 337–348.

Diamond, Peter A (1971), 'A model of price adjustment', *Journal of economic theory* **3**(2), 156–168.

Dubé, Jean-Pierre (2004), 'Multiple discreteness and product differentiation: Demand for carbonated soft drinks', *Marketing Science* **23**(1), 66–81.

Dubé, Jean-Pierre, G Hitsch and Pranav Jindal (2009), Estimating durable goods adoption decisions from stated preference data, Technical report.

Dubé, J.P., G.J. Hitsch and P.E. Rossi (2009), 'Do switching costs make markets less competitive?', *Journal of Marketing Research* **46**(4), 435–445.

Dubé, J.P., G.J. Hitsch and P.E. Rossi (2010), 'State dependence and alternative explanations for consumer inertia', *The RAND Journal of Economics* **41**(3), 417–445.

Ellison, Glenn and Sara Fisher Ellison (2009), ‘Tax sensitivity and home state preferences in internet purchasing’, *American Economic Journal: Economic Policy* pp. 53–71.

Erdem, Tulin and Michael P. Keane (1996), ‘Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets’, *Marketing Science* **15**(1), 1–20.

Erdem, Tülin, Michael P Keane, T Sabri Öncü and Judi Strebel (2005), ‘Learning about computers: An analysis of information search and technology choice’, *Quantitative Marketing and Economics* **3**(3), 207–247.

Forman, Chris, Anindya Ghose and Avi Goldfarb (2006), ‘Competition between local and electronic markets: How the benefit of buying online depends on where you live’, *Management Science* **55**(1), 44–57.

Foster, A.D. and M.R. Rosenzweig (1995), ‘Learning by doing and learning from others: Human capital and technical change in agriculture’, *Journal of Political Economy* pp. 1176–1209.

Gentzkow, Matthew Aaron (2007), ‘Valuing new goods in a model with complementarity: Online newspapers’, *American Economic Review* **97**(3), 713–744.

Goeree, Michelle Sovinsky (2008), ‘Limited information and advertising in the us personal computer industry’, *Econometrica* **76**(5), 1017–1074.

Goolsbee, A (2000), ‘In a world without borders: The impact of taxes on internet commerce.’, *Quarterly Journal of Economics* **115**(2), 561–576.

Gowrisankaran, Gautam and Marc Rysman (2012), ‘Dynamics of consumer demand for new durable goods’, *Journal of Political Economy* **120**(6), 1173–1219.

Hendel, Igal (1999), ‘Estimating multiple-discrete choice models: An application to computerization returns’, *The Review of Economic Studies* **66**(2), 423–446.

Hendel, Igal and Aviv Nevo (2006), ‘Measuring the implications of sales and consumer inventory behavior’, *Econometrica* **74**(6), 1637–1673.

Huang, Peng, Nicholas H. Lurie and Sabyasachi Mitra (2009), 'Searching for experience on the web: An empirical examination of consumer behavior for search and experience goods', *Journal of Marketing* **73**, 55–69.

Jovanovic, B. and Y. Nyarko (1996), 'Learning by doing and the choice of technology', *Econometrica: Journal of the Econometric Society* pp. 1299–1310.

Kasahara, H. and K. Shimotsu (2009), 'Nonparametric identification of finite mixture models of dynamic discrete choices', *Econometrica* **77**(1), 135–175.

Keane, M.P. and K.I. Wolpin (1994), 'The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence', *The Review of Economics and Statistics* pp. 648–672.

Kim, Jaehwan, Greg M Allenby and Peter E Rossi (2002), 'Modeling consumer demand for variety', *Marketing Science* **21**(3), 229–250.

Kim, Jun B, Paulo Albuquerque and Bart J Bronnenberg (2010), 'Online demand under limited consumer search', *Marketing Science* **29**(6), 1001–1023.

Levitt, Steven D, John A List and Chad Syverson (2013), 'Toward an understanding of learning by doing: Evidence from an automobile assembly plant.', *Journal of Political Economy* **121**(4), 643–681.

Lucas Jr, Robert E and Benjamin Moll (2014), 'Knowledge growth and the allocation of time', *Journal of Political Economy* **122**(1).

Magnac, Thierry and David Thesmar (2002), 'Identifying dynamic discrete decision processes', *Econometrica* **70**(2), 801–816.

Melis, Kristina, Katia Campo, Els Breugelmans and Lien Lamey (2014), 'The impact of the multi-channel retail mix on online store choice: Does online experience matter?', *Journal of Retailing* .

Melnikov, Oleg (2000), 'Demand for differentiated durable products: The case of the us computer printer market', *Manuscript. Department of Economics, Yale University* .

Michael, R.T. (1973), 'Education in nonmarket production', *The Journal of Political Economy* pp. 306–327.

Murray, K. and G. Häubl (2005), 'Explaining cognitive lock-in: The role of skill-based habits of use in consumer choice', *Journal of Consumer Research, Vol. 34, No. 1, 2007* .

Nam, Myungwoo, Jing Wang and Angela Y Lee (2012), 'The difference between differences: How expertise affects diagnosticity of attribute alignability', *Journal of Consumer Research* **39**(4), 736–750.

Nelson, Phillip (1970), 'Information and consumer behavior', *The Journal of Political Economy* **78**(2), 311–329.

Nosal, Kathleen (2012), 'Estimating switching costs for medicare advantage plans', *Unpublished manuscript, University of Mannheim* .

Ofek, Elie, Zsolt Katona and Miklos Sarvary (2011), '"bricks and clicks": The impact of product returns on the strategies of multichannel retailers', *Marketing Science* **30**(1), 42–60.

Peterson, Robert A., Sridhar Balasubramanian and Bart J . Bronnenberg (1997), 'Exploring the implications of the internet for consumer marketing', *Journal of the Academy of Marketing Science* **25**(4), 329–346.

Pinna, Fabio and Stephan Seiler (2014), 'Consumer search: Evidence from path-tracking data'.

Prince, Jeffrey T (2007), 'The beginning of online/retail competition and its origins: An application to personal computers', *International Journal of Industrial Organization* **25**(1), 139–156.

Ratchford, Brian T (2001), 'The economics of consumer knowledge', *Journal of Consumer Research* **27**(4), 397–411.

Rust, J. (1987), 'Optimal replacement of gmc bus engines: An empirical model of harold zurcher', *Econometrica: Journal of the Econometric Society* pp. 999–1033.

Seiler, Stephan (2013), 'The impact of search costs on consumer behavior: A dynamic approach', *Quantitative Marketing and Economics* **11**(2), 155–203.

Shaw, Kathryn and Edward Lazear (2008), 'Tenure and output', *Labour Economics* **15**(4), 704–723.

Shugan, Steven M (1980), 'The cost of thinking', *Journal of consumer Research* pp. 99–111.

Simon, Herbert A (1959), 'Theories of decision-making in economics and behavioral science', *The American economic review* pp. 253–283.

Song, Inseong and Pradeep K Chintagunta (2003), 'A micromodel of new product adoption with heterogeneous and forward-looking consumers: Application to the digital camera category', *Quantitative Marketing and Economics* **1**(4), 371–407.

Thomadsen, Raphael (2007), 'Product positioning and competition: The role of location in the fast food industry', *Marketing Science* **26**(6), 792–804.

Van Heerde, Harald J, Peter SH Leeflang and Dick R Wittink (2001), 'Semiparametric analysis to estimate the deal effect curve', *Journal of Marketing Research* **38**(2), 197–215.

Van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel and Philip Hans Franses (2010), 'Retrieving unobserved consideration sets from household panel data', *Journal of Marketing Research* **47**(1), 63–74.

Yao, Song, Carl F Mela, Jeongwen Chiang and Yuxin Chen (2012), 'Determining consumers' discount rates with field studies', *Journal of Marketing Research* **49**(6), 822–841.

Youn, N., I. Song and D. MacLachlan (2008), 'A multi-category approach to modeling consumer preference evolution: The case of sporting goods'.